Probability Distributions for Complex Systems: Adaptive Umbrella Sampling of the Potential Energy

Christian Bartels[†] and Martin Karplus*[,†,‡]

Laboratoire de Chimie Biophysique, Institut Le Bel, Université Louis Pasteur, 4, rue Blaise Pascal, 67000 Strasbourg, France, and Department of Chemistry, Harvard University, Cambridge, Massachusetts 02138

Received: July 14, 1997; In Final Form: October 15, 1997[⊗]

An adaptive umbrella sampling molecular dynamics algorithm based on the potential energy is introduced to reduce the simulation time required for deriving equilibrium properties of molecular systems. The potential energy, in contrast to the commonly used umbrella potentials, is of particular interest for complex systems because it does not depend on assumptions about the geometry nor on a knowledge of the important conformations and transition states that are involved in the equilibration. The method is illustrated by applying it to the threonine dipeptide and to met-Enkephalin. For the threonine dipeptide, potentials of mean force for all dihedral angles of the system are derived from a single 10 ns run. The accuracy of the potentials is confirmed by comparison with previous results. Met-Enkephalin is found to sample several different conformations at 300 K. Three representative structures are used to cluster the dominant conformations. Equilibrium distributions of dihedral angles and the distance between the two termini are calculated. The precision of the results is analyzed, and the utility of the proposed adaptive umbrella sampling method is discussed.

1. Introduction

Empirical potential energy functions are being widely used for the simulation of mesoscopic systems.[1,2] A single energy evaluation in such a simulation, which can contain several thousand or more atoms, involves a sum over a large number of different terms and is computationally expensive. This limits the number of configurations that can be sampled and restricts the kinds of properties that can be derived since ordinary molecular dynamics simulations are restricted to several nanoseconds with present day computers. In particular, it is difficult to evaluate properties which require sampling of different local minima separated by free energy barriers or to determine equilibrium conformations of systems with a large number of degrees of freedom. Consequently, methods are needed for more efficient sampling of conformational space.

In considering the evaluation of the average values of properties at equilibrium, it is convenient to distinguish three different types of conformations. The first set, called significant local minima, are of similar free energies and dominate the equilibrium properties of the system. The second set, called transition states, are rarely sampled in unbiased simulations. Their sampling is crucial for the equilibration of the system and for the correct sampling of different local minima. The barriers are also of interest for understanding the dynamics of the system. The third set consists of the remaining conformations, which are generally of higher energy, have a negligible influence on the equilibrium properties of the system and are unimportant for equilibration. They are referred to as unimportant conformations. An ideal simulation scheme leads to rapid equilibration and predominantly samples the significant local minima and the transition state conformations without generating many unimportant conformations. Such a scheme is difficult to implement for complex systems because it is often necessary to include many degrees of freedom in the sampling procedure. For a peptide, for example, sampling of the significant local minima and transition states of a single dihedral angle will not, in general, lead to the correct distributions for the system as a whole. Also, the conformations must be sampled in such a way that the equilibrium properties of the system can be derived. This means that the entropy, as well as the energy, has to be taken into account in the sampling and that the simulation results can be related to the properties of the system in the appropriate thermodynamic ensemble.

Because of the importance of the sampling problem, a large number of methods to improve upon standard molecular dynamics simulations have been proposed; for a review see ref 3. The methods include high-temperature simulations, sometimes combined with the weighted histogram analysis method,[4−8] umbrella sampling[9] and adaptive umbrella sampling,[6,10,11] potential scaling and potential annealing,[12,13] the introduction of the van der Waals radius as an additional degree of freedom,[14] mean field approximations using multiple copies of the system,[15,16] or smoothing of the potential.[17,18]

Umbrella sampling directs the system into regions of conformational space that are insufficiently sampled in unbiased simulations. This is achieved by adding biasing potentials, called umbrella potentials, to the Hamiltonian of the system. In conventional umbrella sampling, the biasing potentials are defined by trial and error, while in adaptive umbrella sampling uniform sampling is achieved along chosen variables by an iterative procedure. So far adaptive umbrella sampling has been used to enhance sampling of variables directly related to the geometry of important conformations of the system, e.g., distances or dihedral angles. This restricts its application to systems for which the important conformations are known or can be found easily and for which the transitions can be

[†] Université Louis Pasteur.
[‡] Harvard University.
[⊗] Abstract published in *Advance ACS Abstracts,* December 15, 1997.

S1089-5647(97)02280-3 CCC: $15.00 © 1998 American Chemical Society
Published on Web 01/09/1998

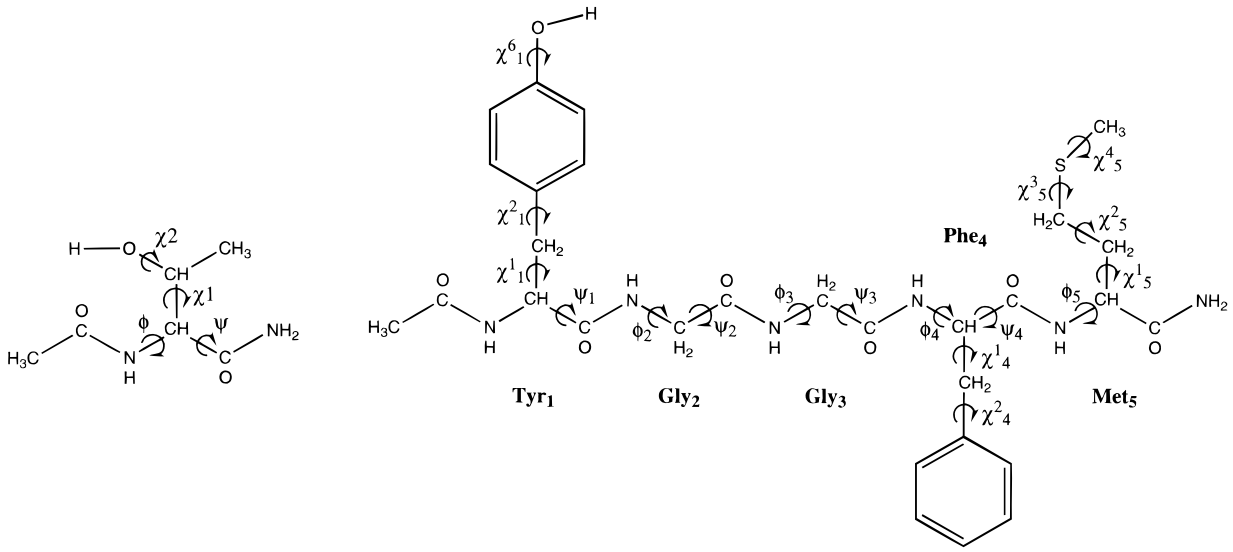

Figure 1. Covalent structure of the threonine dipeptide (A) and met-Enkephalin (B). Dihedral angles mentioned in the text are indicated.

described by a few degrees of freedom. For such systems adaptive umbrella sampling enables sampling of both the transition states and significant local minima.

In the present paper we extend the adaptive umbrella sampling technique to a new choice for the variable along which the umbrella potential is applied. Instead of certain specific coordinates, the variable used is the potential energy of the system. As we show, this provides a general sampling method that can be used to obtain distributions for many degrees of freedom. This is possible since the method samples the significant local minima as well as the transition states without generating too many unimportant conformations.

An important set of closely related sampling methods have been formulated within a Monte Carlo framework. They include entropic sampling,[19] temperature-expanded simulations,[20] and multicanonical sampling.[21,22] These methods, too, give a large increase in sampling of transition states with only a moderate decrease in sampling of significant local minima. They have successfully been applied to systems in vacuum[23−25] and were found to be more efficient than simulated annealing in certain cases.[26]

Section 2 describes the method and relates it to entropic sampling and multicanonical simulations. Details of the calculations are given in section 3. In section 4, the method is validated by applying it to two peptides and deriving certain of their equilibrium properties. They include the distributions for the dihedral angles and distances between atoms, and the density of states with the same potential energy. The derived properties are analyzed with respect to their precision and with respect to results obtained with other simulation techniques. The peptides studied are the threonine dipeptide (*N*-acetyl-L-threoninamide; Figure 1A) and met-Enkephalin[27,28] protected at its N- and C-terminus (acetyl-Tyr-Gly-Gly-Phe-Met-amide; Figure 1B). The threonine dipeptide was chosen since the results could be compared to those of a previous study[6] based on conventional adaptive umbrella sampling. Blocked met-Enkephalin was selected since it is relatively small but is a nontrivial system which is distributed among different conformations at room temperature. Also, it contains no charged groups which would dominate the energetics. The results show that use of the potential energy as the variable for adaptive umbrella sampling achieves correct distributions for these peptides in vacuum. Comparisons are made with the results of other studies of met-Enkephalin; they include a simulation that makes use of a molecular dynamics version of multicanonical sampling,[29] which was published after this work was completed. In a forthcoming publication it is demonstrated that the total potential energy can be used to study small peptides in explicit aqueous solvent.[30] A concluding discussion is given in section 5.

2. Adaptive Umbrella Sampling of the Potential Energy

2.1. Adaptive Umbrella Sampling. Adaptive umbrella sampling[6,10,11,31] of a chosen variable λ leads to uniform sampling of this variable under ideal conditions. A number of simulations i with a modified potential $H_i = H° + U_i(\lambda)$ have to be carried out in which the umbrella potential $U_i(\lambda)$ is added to the Hamiltonian $H°$ of the system. After each simulation the umbrella potential $U_i(\lambda)$ is updated such that uniform sampling of the variable can be expected. This is done based on statistics of the sampling in the previous simulations, $j = 1, 2, ..., i$. For this purpose, the variable λ is partitioned into bins with index k and $l \in (\xi_k, \xi_{k+1})$, where the ξ_k denotes the boundaries between the bins. For each of the simulations, the number of times $n_{j,k}$ in which the system is found in a particular bin is determined. Then an estimate $\tilde{p}^o_k$ for the probability of finding the unperturbed system in a particular bin is calculated by determining a self-consistent solution of the WHAM eqs 1 and 2 (weighted histogram analysis method)[5−8]

$$\tilde{p}^o_k = \frac{\sum_j n_{j,k}}{\sum_j N_j \tilde{f}_j c_{j,k}} \tag{1}$$

$$\tilde{f}_j = \frac{1}{\sum_k c_{j,k} \tilde{p}^o_k} \tag{2}$$

with

$$c_{j,k} = e^{-U_j(\bar{\xi}_k)/RT} \tag{3}$$

$$N_j = \sum_k n_{j,k} \tag{4}$$

In eq 1 the summations are over all simulations and in eqs 2 and 4 over all bins. R is the gas constant, T is the temperature

of the simulations, $\bar{\xi}_k = {}^1/_2\{\xi_k + \xi_{k+1}\}$ is the midpoint of bin k, and $\tilde{f}_j$ is a scaling factor which arises from the condition that the probability of finding the system in any of the bins should be equal to 1. Details concerning the derivation and the self-consistent solution of eqs 1 and 2 are given in ref 6; a maximum likelihood or Bayesian statistical method is used.[32,33] The discrete version of the umbrella potential for the next simulation is related to the calculated estimate $\tilde{p}^{\circ}{}_k$ by

$$U_k = RT \ln \tilde{p}^{\circ}{}_k \tag{5}$$

From this discrete potential, the continuous umbrella potential $U_{i+1}(\lambda)$ for the next simulation is derived by extrapolating it for those bins that were not sampled (see below and ref 6) and fitting it with a set of continuous basis functions.[6] A sum over a series of cosines and sines plus polynomials in the variable λ has been found to work effectively.

In some cases, it may be useful to combine results from simulations at different temperatures to improve the sampling. The WHAM eqs can be used also in this case, if the definition of the $c_{j,k}$ (eq 3) is adapted appropriately and the necessary statistics on the potential energy of the system are collected. In the following derivation, λ is set equal to the potential energy of the system such that the $n_{j,k}$ contains the statistics on the potential energy. To use other variables for the umbrella potential in simulations at different temperatures, the multi-dimensional WHAM eqs[6] would have to be used.

The derivation of the WHAM eqs[6] is based on the relation of the probability $p^{\circ}{}_k$ of finding the system in bin k in an unbiased simulation at T_{ref} and the corresponding probability $p_{j,k}$ in the jth simulation at T_j of the adaptive umbrella sampling run. The probabilities are proportional to the product of the density of states $n(\bar{\xi}_k)$ (see also section 2.4) with the corresponding Boltzmann factors, i.e.,

$$p^{\circ}{}_k \propto n(\bar{\xi}_k) \mathrm{e}^{-\bar{\xi}_k/RT_{\text{ref}}} \tag{6}$$

and

$$p_{j,k} \propto n(\bar{\xi}_k) \mathrm{e}^{-(\bar{\xi}_k + U_i(\bar{\xi}_k))/RT_j} \tag{7}$$

where $\bar{\xi}_k$ is the midpoint of the potential energies in bin k. Combining eqs 6 and 7 by eliminating the density of states and using the normalization condition $\sum_k p_{j,k} = 1$ yields

$$p_{\mathrm{j,k}} = f_j c_{j,k} p^{\circ}{}_k \tag{8}$$

with

$$c_{j,k} = \mathrm{e}^{-U_j(\bar{\xi}_k)/RT_j - \bar{\xi}_k(1/RT_j - 1/RT_{\text{ref}})} \tag{9}$$

and

$$f_j = \frac{1}{\sum_k c_{j,k} p^{\circ}{}_k} \tag{10}$$

2.2. The Potential Energy as the Variable for Adaptive Umbrella Sampling. As stated in the Introduction, an ideal simulation scheme should sample the significant local minima and the transition state conformations without generating many unimportant conformations. A necessary condition for adaptive umbrella sampling to come close to this ideal is that the variable λ is suitable for distinguishing between the three types of conformations.

In the following, it is assumed that properties of the system at a given finite temperature, e.g., 300 K, are being determined. Properties of the system at the given temperature are dominated by conformations with potential energies that lie within a limited range and may be higher than the potential energies of conformations important at lower temperatures (e.g., the global minimum of the system). This is a consequence of the fact that, in general for systems consisting of many particles, the number of conformers with a given potential energy (the density of states, see also section 2.4) is a steeply increasing function. Thus, the product of the exponentially decreasing Boltzmann factor with the density of states, which is proportional to the probability in the canonical ensemble (see below, eq 19) is, in general, a function that is peaked at a potential energy larger than that of the global minimum of the system (Figures 2 and 5);[34,35] in fact, the global minimum may make a negligible contribution to the conformations that are important at the given temperature. It is this aspect of the probability distribution for complex systems that leads to the requirement of a deep, relatively broad minimum region corresponding to the native structure of a protein.[36] Transition states will in general correspond to conformations with higher potential energies. This is, however, not always the case. For example, for a protein at its denaturation temperature, conformations with potential energies between those typical for the native and those typical for the denatured state may be transition states due to the entropic (density of states) contribution to the free energy.[37] The unimportant conformations will in general have potential energies that are larger than those of the transition states or smaller than those of the significant local minima. It cannot be ruled out, however, that unimportant conformations exist that are far away in conformational space from any significant local minima and have potential energies similar to those of the transition states.

It follows that the potential energy should in many (but not all) cases be suitable to distinguish between the three types of conformations. Therefore, we set the variable λ equal to the potential energy V of the system. The Hamiltonian for the ith simulation is then given by

$$H_i = H^{\circ} + U_i(V) = K + V + U_i(V) \tag{11}$$

where K and V denote the kinetic and true potential energy of the system, respectively, H° is the Hamiltonian of the unperturbed system, and $U_i(V)$ is the umbrella potential applied in the ith simulation. The gradient of the Hamiltonian with respect to the Cartesian coordinates of atom n is required in molecular dynamics simulations; it is simply given by

$$\nabla_n H_i = \left(1 + \frac{\partial U_i}{\partial V}\right) \nabla_n V \tag{12}$$

2.3. Calculation of Observables. Observables of a system at selected conditions can be calculated from simulations of the system at different conditions (e.g., at a different temperature or with an umbrella potential) by weighting each conformation by an appropriate factor.[38] For example, the observable $\langle A \rangle_{T_2}$ of a system at temperature T_2 can be calculated from the values A_t found for the observable in a simulation at temperature T_1 using the weighting factors $\exp((\beta_1 - \beta_2)V_t)$. The resulting expression is

$$\langle A\rangle_{T_2} = \frac{\sum_t A_t \exp((\beta_1 - \beta_2)V_t)}{\sum_t \exp((\beta_1 - \beta_2)V_t)} \tag{13}$$

Here the index t runs over the time steps of the simulation at T_1, V_t denotes the potential energy at time step t, and $\beta_1 = 1/RT_1$ and $\beta_2 = 1/RT_2$ with R being the gas constant. Similarly, values of observables for the correct Hamiltonian can be derived from an adaptive umbrella sampling run. For this purpose, each conformation t_i of simulation i has to be weighted by a factor $\alpha_{K(t_i)}$ which depends on the bin $K(t_i)$ in which the system is found at this time step. The factor is given by[6]

$$\alpha_{K(t_i)} = \frac{1}{\sum_j N_j \tilde{f}_j c_{j,K(t_i)}} \tag{14}$$

where N_j, $\tilde{f}_j$ and $c_{j,K(t_i)}$ are obtained from eqs 2−4. The observable $\langle A\rangle$ is the weighted average of its value A_{t_i} in each of the conformations, i.e.,

$$\langle A\rangle = \frac{\sum_{i,t_i} A_{t_i}\alpha_{K(t_i)}}{\sum_{i,t_i} \alpha_{K(t_i)}} \tag{15}$$

The summations are over all the conformations t_i of all simulations i. To obtain the value of the observable $\langle A\rangle_{T_2}$ at a temperature T_2, which is different from the temperature T_1 used for the umbrella sampling simulations, each conformation has to be weighted as in eq 13 by the additional factor $\exp((\beta_1 - \beta_2)V_{t_i})$, i.e.,

$$\langle A\rangle_{T_2} = \frac{\sum_{i,t_i} A_{t_i}\alpha_{K(t_i)} \exp((\beta_1 - \beta_2)V_{t_i})}{\sum_{i,t_i} \alpha_{K(t_i)} \exp((\beta_1 - \beta_2)V_{t_i})} \tag{16}$$

where V_{t_i} denotes the potential energy of conformation t_i in simulations i.

In section 4, we use eqs 15 and 16 to calculate distributions or potentials of mean force for selected variables; i.e., dihedral angles and intramolecular distances. This is done by partitioning the chosen variable into a number of appropriately sized bins. For each bin l, the probability, p_l, that the system samples this bin is calculated. Let $\delta^l_{t_i}$ denote a function that is 1 if the system at time step t_i is in bin l, and 0 otherwise. Then the probability for the system to be in this bin is estimated according to eq 15 by

$$p_l = \frac{\sum_{i,t_i} \alpha_{K(t_i)}\delta^l_{t_i}}{\sum_{i,t_i} \alpha_{K(t_i)}} \tag{17}$$

From this probability the potential of mean force E_l for this bin is given, up to an additive constant which is the same for all bins, by

$$E_l = -RT \ln(p_l) \tag{18}$$

2.4. Relation between the Umbrella Potential and the Density of States. The probability density $p(H^\circ,V,\beta)$ for conformations with potential energy V to occur in a simulation with Hamiltonian H° at the inverse temperature $\beta = 1/RT$ is proportional to

$$p(H^\circ,V,\beta) \propto n(V) \exp(-\beta V) \tag{19}$$

where $n(V)$ denotes the density of states with potential energy V. In a simulation with the umbrella potential $U_i(V)$ added to the Hamiltonian (eq 11), the corresponding probability density $p(H_i,V,\beta)$ for conformations with potential energy V is proportional to $n(V) \exp(-\beta U_i(V)) \exp(-\beta V)$. The umbrella potential function $U_i(V)$ is assumed to be such that all potential energies are visited the same number of times, i.e.,

$$p(H_i,V,\beta) \propto n(V) \exp(-\beta U_i(V)) \exp(-\beta V) = \text{const} \tag{20}$$

Solving this equation for $n(V)$ yields

$$n(V) \propto \exp(\beta\{U_i(V) + V\}) \tag{21}$$

which determines the density of states as a function of the potential energy V up to a constant multiplier. Given the density of states, the configuration integral for a system in the canonical ensemble can be obtained as

$$Z = \int \mathrm{d}V\, n(V) \exp(-\beta V) \propto \int \mathrm{d}V \exp(\beta U_i(V)) \tag{22}$$

with the integration performed over all potential energies V. Once the configuration integral is known, other thermodynamic quantities can be obtained by use of standard statistical mechanical relationships.[35]

2.5. Range of Sampling. There are several reasons to restrict the range of potential energies that are sampled during the simulations. First, the calculation can be restricted to make them more efficient because it is not necessary to sample all potential energies, since, in general, the important local minima and transition state conformations at a given temperature have energies that lie within a certain range (see section 2.2). Second, in an energy sampling run at temperature T_1, it is inefficient to sample conformations with very low potential energies corresponding to temperatures $T_2 \ll T_1$ approaching zero. To sample such low-energy conformations, umbrella potentials would be required for which $\partial U_i(V)/\partial V$ becomes very large (see below, eq 27). Such an umbrella potential would lead to large forces (eq 12) and would necessitate very small integration time steps to obtain proper integration of the equations of motion. Further, the implementation of the method is simplified if the range of potential energies for which statistics are acquired is specified prior to starting the run and if sampling is restricted to within this specified range. Otherwise, a scheme would be necessary to define the umbrella potential outside of this region and to make sure that the system will return within reasonable time to the region of interest.

In the current implementation, a scheme (see below) is used to restrict the sampling of the potential energies which requires that minimal and maximal temperatures, $T_{\min}$ and $T_{\max}$, and minimal and maximal potential energies, $V_{\min}$ and $V_{\max}$, are specified prior to starting the run. In general, $T_{\min}$ should be set equal to a temperature at or slightly below the temperature for which one wants to derive equilibrium properties, and $T_{\max}$ to a temperature sufficiently high to ensure frequent transitions between different local minima. The range $V_{\min}$ to $V_{\max}$ is then chosen such that it contains the most important potential energies of canonical ensembles at temperatures between $T_{\min}$ and $T_{\max}$.

Approximate values for V_{min} and V_{max} can easily be obtained from short test simulations at T_{min} and T_{max}.

The potential energy limits, V_{min} and V_{max}, are used to define the range of potential energies for which to acquire statistics and to ensure that the system remains within this range. To achieve this, statistics for potential energies are acquired in the range

$$V'_{min} = V_{min} - {}^1/_2(V_{max} - V_{min}) \quad \text{to} \tag{23}$$

$$V'_{max} = V_{max} + {}^1/_2(V_{max} - V_{min}) \tag{24}$$

and extensive sampling outside of the limits V_{min} and V_{max} is prevented by removing the forces due to the bias outside of the limits, i.e., by enforcing

$$\partial U_i(V)/\partial V = 0 \quad \text{for} \quad V > V_{max} \quad \text{and for} \quad V < V_{min} \tag{25}$$

It has been found that to prevent umbrella potentials for which $\partial U_i(V)/\partial V$ becomes too large, accurate estimates for V_{min} would be required. Accurate estimate are, however, difficult to obtain since it would be necessary that the system is well equilibrated, i.e., a full adaptive sampling run would be necessary. This problem can be circumvented by using the temperature limits, T_{min} and T_{max}, to restrict sampling to the conformations that would be sampled in unbiased simulations at temperatures between T_{min} and T_{max}. To do this, a relation is needed between the umbrella potential at a given potential energy and the temperature of the canonical ensemble in which the given potential energy is likely to be sampled. Given the umbrella potential $U_i(V)$ and the density of states $n(V)$ determined in an adaptive umbrella sampling run at temperature T_1, the temperature T_2 of a canonical ensemble can be related to the most likely potential energy V_2 in this ensemble: The probability density $p(H^\circ,V,\beta_2)$ for conformations with potential energy V to occur in an unbiased simulation with Hamiltonian H° at the inverse temperature $\beta_2 = 1/RT_2$ is proportional to $n(V)$ $\exp(-\beta_2 V)$ (eq 19). The function $p(H^\circ,V,\beta_2)$ has a maximum at

$$\frac{\partial p(H^\circ,V,\beta_2)}{\partial V} = 0$$

$$= \frac{\partial \ln p(H^\circ,V,\beta_2)}{\partial V} p(H^\circ,V,\beta_2) \tag{26}$$

$$= \frac{\partial \ln n(V)}{\partial V} - \beta_2$$

Using eq 21 to substitute for $n(V)$ yields

$$\frac{\partial(\beta_1 U_i(V) + \beta_1 V)}{\partial V} - \beta_2 = \beta_1 \frac{\partial U_i(V)}{\partial V} + \beta_1 - \beta_2 = 0$$

which upon substitution of the inverse temperatures and rearrangement results in

$$\left.\frac{\partial U_i(V)}{\partial V}\right|_{V=V_2} = \frac{T_1}{T_2} - 1 \tag{27}$$

From eq 27 it is seen that for canonical ensembles at temperatures outside the range of T_{min} to T_{max}, the derivative $\partial U_i(V)/\partial V$ is larger than $(T_1/T_{min}) - 1$ or smaller than $(T_1/T_{max}) - 1$, respectively (T_1 denotes the simulation temperature). Therefore, the temperature limits can be enforced by modifying the umbrella potential such that

$$\frac{T_1}{T_{min}} - 1 > \frac{\partial U_i(V)}{\partial V} > \frac{T_1}{T_{max}} - 1 \tag{28}$$

Limiting the sampling with eq 28 explicitly restricts $\partial U_i(V)/\partial V$ to within a certain range and therefore provides a convenient and secure way to prevent $\partial U_i(V)/\partial V$ from becoming too large. The considerations involved in restricting the sampling are illustrated in Figure 2.

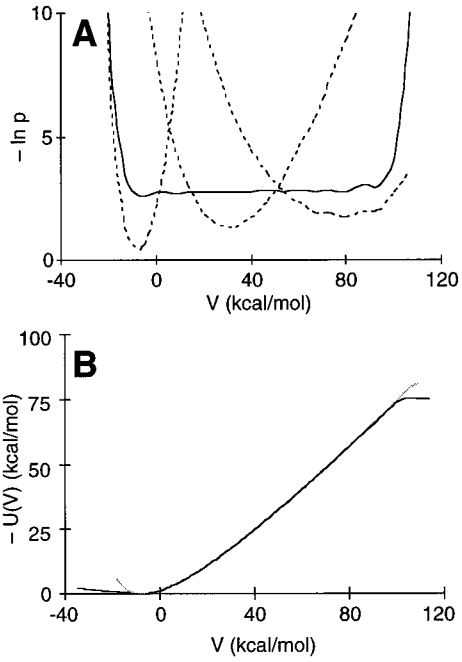

Figure 2. (A) Probability distributions of the potential energy of the threonine dipeptide in the potential energy adaptive umbrella sampling run and in canonical ensembles at 300, 1000, and 2000 K (see text). Sampling in the potential energy sampling run extended up to 108 kcal/mol; no probabilities are plotted for potential energies larger than 108 kcal/mol. The canonical distributions were calculated with eq 17. (B) Final umbrella potential of the 10 ns energy sampling run. Gray line, discrete umbrella potential (eq 5); black line, continuous umbrella potential obtained after applying eqs 25 and 28 to restrict sampling to the desired range of potential energies.

Using both temperature and potential energy limits to restrict sampling is somewhat redundant, but simplifies the implementation. The potential energy limits are useful since otherwise more complicated schemes would be needed to define the range for which statistics are acquired and to keep the system within this range; the temperature limits are useful since otherwise more complicated schemes (e.g., a dynamic adaptation of the potential energy limits) would be required to prevent $\partial U_i(V)/\partial V$ from becoming too large or a scheme would have to be introduced to reduce the integration time step if $\partial U_i(V)/\partial(V)$ becomes large. By doing the latter it would be possible to find the global minimum structure, which is, of course, the single conformation that makes the largest contribution at any finite temperature (see also section 2.2).

2.6. Extrapolation of the Umbrella Potential. In adaptive umbrella simulations many bins in the interesting regions will not be sampled in the initial stages. As we showed in the earlier paper on adaptive sampling for dihedral angles, an extrapolation scheme for the probability of those bins can significantly improve the rate of convergence. An extrapolation scheme should direct the system into regions that were not yet sampled but should not introduce additional local minima. The perfor-

mance of an extrapolation scheme depends on the kind of umbrella potential that has to be extrapolated. The extrapolation scheme proposed previously for dihedral angle variables[6] resulted in rather slow exploration of new bins in the present case (results not shown). In the present applications we first tested linear extrapolation as an alternative. Let k_1 and k_2 be the indices of the first and last bin, respectively, of the bins that were already sampled. Then the umbrella potential is extrapolated for all the bins $k' < k_1$ according to

$$U_{k'} = U_{k_1} + (k_1 - k')(U_{k_1} - U_{k_1+1}) \tag{29}$$

and for all bins $k'' > k_2$ according to

$$U_{k''} = U_{k_2} + (k'' - k_2)(U_{k_2} - U_{k_2-1}) \tag{30}$$

This linear extrapolation introduced additional local minima in some cases (results not shown), and the system became trapped in these minima for a number of simulations. To avoid this problem, a "damped" linear extrapolation,

$$U_{k'} = U_{k_1} + (U_{k_1} - U_{k_1+1})\frac{1 - q^{k_1-k'}}{1 - q} \tag{31}$$

$$U_{k''} = U_{k_2} + (U_{k_2} - Uk_{_2-1})\frac{1 - q^{k''-k_2}}{1 - q} \tag{32}$$

with $0 < q < 1$, was introduced. In all the cases studied in this paper good convergence of the umbrella potential was achieved with $q = 0.8$ (an example is shown Figure 7B).

The extrapolation scheme assumes that the bins that were already sampled form a continuous set. This is usually true. However when gaps do occur, they are filled prior to the application of eqs 31 and 32 using linear extrapolation between the two bins at the border of the gap.

2.7. Comparison with Alternative Methods. With a function of the potential energy as the adaptive umbrella potential the present molecular dynamics method is closely related to entropic sampling or multicanonical simulations. The latter have been implemented primarily in a Monte Carlo framework. In these methods, as in the presented one, the probability of sampling particular conformations is biased such that all potential energies in the range of interest for the system are sampled equally. By so doing, the density of states, $n(V)$, is determined implicitly since it is the product of $n(V)$ and the Boltzmann factor that gives the probability of configurations with a given potential energy (see eq 19).

In Monte Carlo simulations, transition probabilities between two conformations are directly related to the ratio of their Boltzmann factors $e^{V_t/RT}$ (R being the gas constant, T the simulation temperature, and V_t the potential energy of conformation t). It is therefore convenient to implement a biased sampling method to obtain a uniform distribution by replacing the Boltzmann factors by $e^{-W(V_t)}$, where the biasing function $W(V_t)$ is determined iteratively to obtain uniform sampling of the potential energies in a selected range. This is what is done in entropic sampling; the term entropic sampling is used because $RW(V_t)$ is the entropy of the states with potential energy V_t. In multicanonical Monte Carlo simulations the bias is introduced by replacing the Boltzmann factors with $e^{-\beta(V_t)V_t+\alpha(V_t)}$, where the functions $\alpha(V_t)$ and $\beta(V_t)$ are the biasing functions. Again they are determined iteratively to obtain uniform sampling of the potential energies; the term multicanonical indicates that the resulting sampling can be viewed as an average over a set of canonical ensembles at different temperatures. In molecular dynamics simulations, sampling is directly related to the gradient of the Hamiltonian with respect to the spatial coordinates, i.e., the forces. The bias is therefore most easily expressed as a modification of the Hamiltonian and not of the Boltzmann factors, i.e., as an umbrella potential $U_i(V_t)$ which is added to the Hamiltonian of the system. This gives $e^{-(V_t+U_i(V_t))/RT}$ for the modified Boltzmann factor. These three formulations (entropic sampling, multicanonical Monte Carlo simulations, and molecular dynamics simulations with an umbrella potential) differ only in how convenient they are for certain applications. They use the same type of bias (see also ref 39), and they are related to each other by

$$W(V_t) = \beta(V_t)V_t - \alpha(V_t) = \frac{1}{RT}(V_t + U_i(V_t)) \tag{33}$$

The bias is determined iteratively in a series of simulations. The methods differ in the way the bias is parametrized and in the way it is updated after each simulation.

While this work was in progress, Nakajima et al.[29] and Hansmann et al.[40] used the entropic formulation of the adaptive Monte Carlo method as a basis for developing a molecular dynamics method. For all of the methods the final resulting bias should be the same, in principle, although the formulation of Nakajima et al. is somewhat more complicated than that used here because it was not expressed in terms of a modification of the Hamiltonian of the system. Also, there are important differences between the formulations of Nakajima et al.[29] and of Hansmann et al.[40] and the present implementation in the way that the umbrella potentials are refined. To derive the umbrella potential for the ith simulation, Nakajima et al. and Hansmann et al. use exclusively data from the previous simulation, $i - 1$ (refs 21, 26, 40 and eq 5 of ref 29). This leads to a problem when choosing the frequency for updating the umbrella potential. To obtain precise estimates of the umbrella potential, long simulations (few updates) are required. To rapidly obtain first estimates of the umbrella potential and thereby to extend rapidly the range of potential energies which are sampled during the early stages of a run, frequent updates (short simulations) are most effective. With the present method, the umbrella potential for the ith simulation is determined based on the data from all previous simulations. Thus, the umbrella potential can be updated frequently to ensure rapid exploration of potential energies at the beginning of a run; properties derived from the run will be precise since the estimates will be based on all the simulations of the run. For the characterization of met-Enkephalin, Nakajima et al.[29] had to perform 2 ns of simulation to obtain a good estimate of the umbrella potential followed by a 5 ns production simulation. Hansmann et al.[40] investigated the same system, but they do not report how many molecular dynamics integration steps were required to obtain a good estimate of the umbrella potential; their production simulation consisted of 7 600 000 molecular dynamics steps. By contrast, we find in section 4 that the present method requires only 2 ns (2 000 000 molecular dynamics steps) to derive estimates of the equilibrium properties of the same system. Clearly, such a comparison is preliminary as long as statistical uncertainties of the estimates cannot be compared; no such data were given by either Nakajima et al.[29] or Hansmann et al.[40]

Nakajima et al.[29] point out that the Nosé−Hoover thermostat[41,42] might be inappropriate for temperature control in multicanonical molecular dynamics simulations. For the examples presented in section 4 we used the Nosé−Hoover thermostat with a coupling constant of 20 kcal ps^2 for temperature control and did not detect any problems. We also

performed adaptive umbrella sampling runs of met-Enkephalin at two different temperatures (300 and 1000 K). If temperature control were a problem, one would expect these two runs to result in different umbrella potentials or different equilibrium properties of the system. There were no significant differences in the results, indicating that any artifacts introduced by the Nosé−Hoover thermostat are small, at least for the present systems.

3. Details of the Calculations

Molecular dynamics simulations were performed using the program CHARMM[43] with the all-hydrogen parameter set PARAM22.[44,45] The time step for the integrator was set to 1 fs, and the SHAKE algorithm[46] was used to fix the lengths of bonds involving hydrogen atoms.

3.1. Threonine Dipeptide. For the isolated threonine dipeptide (Figure 1A), a distance-dependent dielectric was used and all nonbonded interactions were truncated at 9 Å with a shift function.[43] Starting from an extended structure with the PARAM22 default values for the internal coordinates (i.e., $\phi = -60°$, $\psi = 0°$, $\chi1 = 180°$, $\chi2 = -179°$), the dipeptide was prepared by 50 steps of steepest descent minimization to remove bad contacts, 3000 steps of dynamics at 1000 K, and then cooling the system down to 303 K in 2000 additional steps of dynamics; Langevin dynamics with a friction coefficient of 6 ps^{-1} was used. The actual adaptive umbrella sampling runs consisted of alternating periods of equilibration and production; a Nosé−Hoover thermostat[41,42] with a coupling constant of 20 kcal ps^2 was employed. Coordinates were saved every 10 time steps.

Two adaptive umbrella sampling runs with the total potential energy as the variable for umbrella sampling were performed at 303 K. First, a short run consisting of 100 updates of the umbrella potential was performed, and then a long run consisting of 1000 updates. Before each update, 2000 molecular dynamics steps were performed for equilibration, followed by 8000 molecular dynamics steps during which statistics on the potential energy were acquired. To derive the properties of the system at 300 K, T_{min} (eq 28) was set equal to 280 K, while T_{max} (eq 28) was set equal to 2000 K. The choice of T_{min} only slightly below 300 K was used since conformations at lower temperatures are unimportant to derive properties at 300 K but would necessitate a small integration time step (section 2.5). The value for T_{max} was chosen relatively high since it was observed that transitions between different side chain rotamers become frequent on a ps time scale only at simulation temperatures above 1000 K (results not shown). Two short unbiased simulations at 280 and 2000 K, respectively, showed that in this temperature interval the system samples potential energies within the range −50 to 100 kcal/mol. Consequently, V_{min} and V_{max} (eq 25) were set to these values. Statistics were collected for potential energies in the energy range V'_{min} to V'_{max} (eqs 23 and 24) using 200 bins of equal size. The umbrella potential was represented as a linear combination of 40 trigonometric functions

$$\sin\left(\frac{2\pi nV}{V'_{max} - V'_{min}}\right) \quad \text{and} \quad \cos\left(\frac{2\pi nV}{V'_{max} - V'_{min}}\right) \qquad (34)$$

with $n = 1, 2, ..., 20$, and four polynomial functions

$$\left(\frac{V - V'_{min}}{V'_{max} - V'_{min}}\right)^n \qquad (35)$$

TABLE 1: Sampling Probabilities for the Dihedral Angle $\chi1$ of the Threonine Dipeptide in Canonical Ensembles at Different Temperatures and in the Adaptive Umbrella Sampling Run

$\chi1$ (deg)	300 K[a]	1000 K[a]	2000 K[a]	avg[b]	energy umb[c]
−180 to −160	9.5E−03[d]	1.0E−01	8.9E−02	6.6E−02	8.0E−02
−160 to −140	7.7E−04	4.0E−02	5.9E−02	3.3E−02	3.8E−02
−140 to −120	3.1E−06	5.9E−03	2.3E−02	9.7E−03	1.1E−02
−120 to −100	2.2E−09	3.1E−03	1.5E−02	6.0E−03	6.4E−03
−100 to −80	2.4E−05	1.3E−02	2.5E−02	1.3E−02	1.5E−02
−80 to −60	1.1E−03	6.1E−02	6.4E−02	4.2E−02	5.1E−02
−60 to −40	1.0E−02	1.1E−01	1.0E−01	7.5E−02	9.2E−02
−40 to −20	6.6E−03	7.4E−02	8.2E−02	5.4E−02	6.5E−02
−20 to 0	3.0E−04	3.2E−02	5.3E−02	2.8E−02	3.3E−02
0 to 20	3.9E−04	3.3E−02	5.1E−02	2.8E−02	3.3E−02
20 to 40	4.4E−02	1.1E−01	8.2E−02	7.9E−02	9.3E−02
40 to 60	6.5E−01	2.1E−01	1.2E−01	3.3E−01	2.6E−01
60 to 80	2.7E−01	1.2E−01	9.1E−02	1.6E−01	1.3E−01
80 to 100	3.0E−03	2.0E−02	4.3E−02	2.2E−02	2.3E−02
100 to 120	1.1E−05	2.7E−03	1.2E−02	5.0E−03	5.6E−03
120 to 140	8.8E−11	1.7E−03	8.9E−03	3.5E−03	4.0E−03
140 to 160	1.5E−05	9.2E−03	2.0E−02	9.7E−03	1.1E−02
160 to180	3.4E−03	6.0E−02	6.1E−02	4.1E−02	4.9E−02

[a] Probability for the $\chi1$ values given in the first column in the canonical ensemble at the specified temperature. [b] Average Probability of the probabilities in the three canonical ensembles. [c] Probability for the $\chi1$ values given in the first column in the adaptive umbrella sampling run. [d] Read as 9.5×10^{-3}.

with $n = 0, 1, 2, 3$. To ensure that the fit gives a unique solution and no wiggles are introduced into the potential, the number of functions used to represent the umbrella potential must be smaller than the number of bins. To ensure that the umbrella potential is well represented, the number of functions has to be sufficiently large. With the present choice of 40 trigonometric and 4 polynomial functions, differences between the discrete and continuous umbrella potential are smaller than 0.1 kcal/mol (see also Figure 2B). In the simulations we did not detect any problems related to the number of functions that were used to represent the umbrella potential. Therefore, we did not experiment with changing the number of functions. Fewer functions are likely to be adequate (see also Figure 8); as remarked by one of the referees and as mentioned in the discussion of Figure 5A, small inaccuracies in the umbrella potential do not affect the expectation values derived from the simulations.

To compare adaptive umbrella sampling of the potential energy with established methods, we use the potential of mean force for the dihedral angle $\chi1$ which we have determined with standard adaptive umbrella sampling in earlier work.[6] In the latter the dihedral $\chi1$ was used as the variable λ along which the umbrella potential was applied. For these calculations, one hundred iterations ($i = 1, 2, ..., 100$) of the adaptive umbrella sampling algorithm were performed; each consisted of 1000 steps of equilibration and 9000 steps for the acquisition of the statistics, followed by an update of the umbrella potential. Statistics for $\chi1$ were collected using 180 bins of two degrees each. Twelve trigonometric functions, $\sin(n\chi1)$ and $\cos(n\chi1)$ ($n = 1, 2, ..., 6$), plus a constant were used as basis functions to represent the continuous umbrella potential. Details are given in ref 6.

3.2. Met-Enkephalin. For the isolated met-Enkephalin, acetyl-Tyr-Gly-Gly-Phe-Met-amide (Figure 1B), a distance-dependent dielectric was used and all nonbonded interactions were truncated at 14.5 Å with a shift function.[43] Starting from an extended structure with the PARAM22 default values for the internal coordinates (dihedral angles are listed in Table 2),

TABLE 2: Dihedral Angles of the Three Representative Structures in Figure 11, the Selected Low-Energy Conformer, and the Minimum Energy Conformer from Ref 23

	start[a]	structure A	structure B	structure C	low energy[b]	ref 23
ψ_1	180	134	122	112	132 (128)	151
ϕ_2	180	84	−139	−139	−158 (−167)	−158
ψ_2	180	−57	53	66	73 (77)	71
ϕ_3	180	−68	88	58	85 (94)	64
ψ_3	180	−46	−72	−46	−61 (−71)	−94
ϕ_4	180	−103	−94	−125	−108 (−110)	−83
ψ_4	180	−64	−63	−141	175 (172)	−30
ϕ_5	180	−95	−87	−87	−97 (−91)	−80
χ_1^1	180	−67	−168	−170	−175 (177)	179
χ_1^2	90	103	33	−112	−113 (−101)	−111
χ_1^6	175	−10	−42	120	143 (157)	149
χ_4^1	180	−65	60	−168	−162 (−154)	180
χ_4^2	90	−114	104	46	67 (55)	79
χ_5^1	180	−70	55	−76	−173 (164)	−66
χ_5^2	180	−176	−177	68	62 (69)	−176
χ_5^3	180	−104	160	−175	73 (76)	−179
χ_5^4	180	−71	−144	47	−64 (−60)	−61

[a] Extended starting conformer (before equilibration and adaptive umbrella sampling). [b] Selected low-energy conformer (see text) after energy minimization. The values in parentheses are the dihedral angles of the conformer before energy minimization.

met-Enkephalin was prepared by 500 steps of steepest descent minimization, 5000 steps of dynamics at 1000 K, and a further 5000 steps of dynamics at the temperature of the umbrella sampling run (300 or 1000 K, respectively; see below); Langevin dynamics with a friction coefficient of 3 ps^{-1} was used.

In principle, runs of umbrella sampling of the potential energy at different temperatures should give the same results if full convergence has been achieved. To test this, two runs were performed at 1000 K with 100 updates of the umbrella potential and one run at 300 K with 200 updates. Before each update 4000 molecular dynamics steps for equilibration were performed; they were followed by 16000 molecular dynamics steps (16 ps) during which statistics on the potential energy were acquired. A Nosé−Hoover thermostat[41,42] with a coupling constant of 20 kcal ps^2 was employed. Coordinates were saved every 20 time steps. T_{min} and T_{max} (eq 28) were set to 280 and 1500 K. A lower value than for the threonine dipeptide was chosen for T_{max} since at temperatures above 1500 K unimportant conformations with cis peptide bonds are frequently sampled. Based on short test simulations at 280 and 1500 K, V_{min} and V_{max} were set to −500 and 500 kcal/mol, respectively. Statistics were collected for potential energies in the energy range V'_{min} to V'_{max} (eqs 23 and 24) using 400 bins of equal size. The umbrella potential was represented as a linear combination of 80 trigonometric functions (eq 34 with $n = 1, 2, ..., 40$) and three polynomial functions (eq 35 with $n = 0, 1, 2$).

4. Results

4.1. Threonine Dipeptide. Figure 2A shows the probability distribution of the potential energy in the long, 10 ns, potential energy adaptive umbrella sampling run and compares it to the probability distributions of the canonical ensembles at 300, 1000, and 2000 K. The probability distribution of the adaptive umbrella sampling run is flat for the most probable potential energies in canonical ensembles at temperatures between T_{min} (280 K) and T_{max} (2000 K), i.e., between −9.5 and 80 kcal/mol. The probability distribution of each of the canonical ensembles at temperatures between 280 and 2000 K is much narrower. Together, the canonical ensembles cover the same range of potential energies as the adaptive umbrella sampling run.

In Figure 2B, the discrete umbrella potential (eq 5) at the end of the energy sampling run is shown and compared to the corresponding continuous umbrella potential. The latter is a linear combination of basis functions (eqs 34 and 35) which is obtained by a least-squares fit to the discrete potential after applying eqs 25 and 28 to restrict sampling to the desired range of potential energies. By use of eq 25, the continuous umbrella potential is flat for potential energies larger than 100 kcal/mol (Figure 2B) and smaller than −50 kcal/mol (not shown), by use of eq 28, the slope of the continuous umbrella potential is confined to within the limits (303 K/280 K) − 1 = 0.082 and (303 K/2000 K) − 1 = −0.85. The slope of the umbrella potential exceeds these limits only for potential energies smaller than −9.5 kcal/mol or larger than 80 kcal/mol. The changes due to the use of eq 28 are readily visible in Figure 2B ($-U(V)$ is plotted) for potential energies smaller than 9.5 kcal/mol; for potential energies larger than 80 kcal/mol, the differences are too small to be seen, i.e., the difference at $V = 100$ kcal/mol is only 0.7 kcal/mol. Due to all these modifications (eqs 25 and 28), the probability distribution of the umbrella sampling run decreases at low potential energies in the same way as the probability distribution of the canonical ensemble at 280 K. At potential energies between 80 and 100 kcal/mol, the probability distribution of the umbrella sampling run decreases in the same way as the probability distribution of the canonical ensemble at 2000 K. At potential energies larger than 100 kcal/mol, the probability distribution of the umbrella sampling run decreases faster than the distribution of the canonical ensemble at 2000 K. These modifications save some simulation time and have negligible effects on properties calculated for the system at room temperature or 1000 K, since potential energies important at these temperatures are sampled appropriately (Figure 2A).

Sampling along degrees of freedom other than the potential energy is illustrated in Figure 3. In Figure 3A, the probability of sampling particular values of $\chi 1$ is plotted for canonical ensembles at 300, 1000, and 2000 K, and for the energy sampling run. The averages of the probabilities for the three canonical ensembles are similar to the probabilities for the energy sampling run (Table 1). This illustrates the fact that the sampling probabilities of the energy sampling run can be expressed as a weighted average of the sampling probabilities of the canonical ensembles at temperatures between 280 and 2000 K. Of the three canonical ensembles for which data is shown in Figure 3A, the canonical ensemble at the average temperature, 1000 K, has a sampling probability distribution most similar to the distribution of the energy sampling run. The distributions at 300 and 2000 K are steeper or smoother, respectively.

Uniform sampling of the potential energies does not imply that the Boltzmann factors are the same for conformations with different potential energies (this would be the case for a simulation at infinite temperature), but that the product of the density of states with the Boltzmann factor is constant for different potential energies (eq 20). Since the density of states generally increases with increasing potential energy, it follows that the probability of sampling any particular high-energy conformer is smaller than the probability of sampling any particular low-energy conformer. This is readily visible in Figure 3A: $\chi 1$ values corresponding to high-energy eclipsed

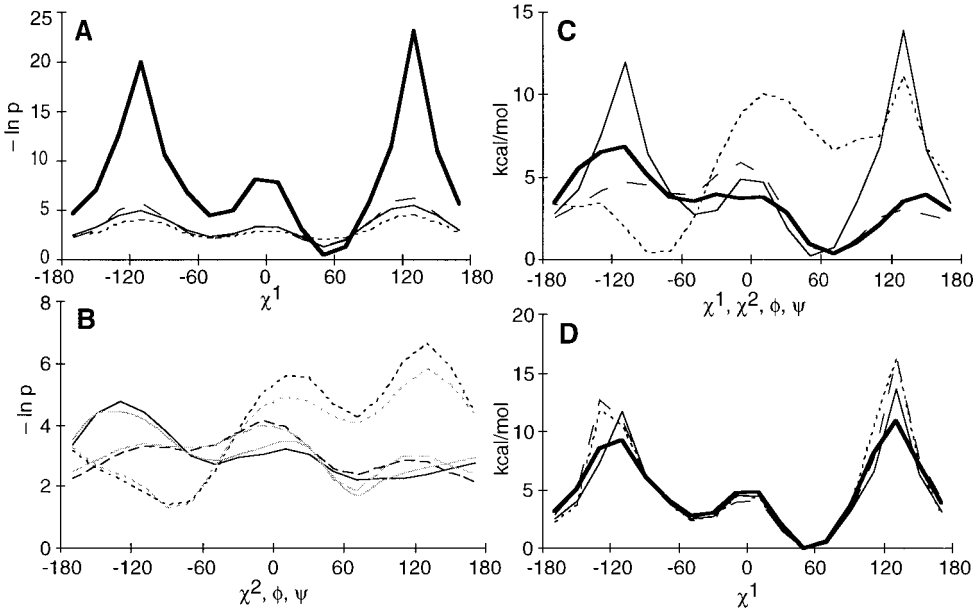

Figure 3. Potentials of mean force and probability of sampling particular values of dihedral angles of the threonine dipeptide. (A) Negative logarithm of the sampling probabilities of the dihedral angle $\chi 1$ in canonical ensembles at 300 K (thick line), at 1000 K (long dashes), and at 2000 K (short dashes), and in the energy sampling run (thin continuous line). (B) Negative logarithm of the sampling probabilities of the dihedral angles $\chi 2$ (long dashes), ϕ (short dashes), and ψ (continuous line) in the canonical ensemble at 1000 K (black lines) and in the energy sampling run (gray lines). (C) Potentials of mean force at 300 K for the dihedral angles $\chi 1$ (continuous thin line), $\chi 2$ (long dashes), ϕ (short dashes), and ψ (continuous thick line) derived from the 10 ns energy sampling run. (D) Potentials of mean force of the dihedral angle $\chi 1$ derived from different energy sampling runs (thin continuous line, 10 ns run; short dashes, first ns of the 10 ns run; long dashes, a different 1 ns run) and from the dihedral angle adaptive umbrella sampling run (thick line); see ref 6.

conformations are always less likely to be sampled than those corresponding to low-energy staggered conformations. For example, in the energy sampling run, $\chi 1$ values between 40° and 60° are ~64 times more often sampled than $\chi 1$ values between 120° and 140°; for the canonical ensemble at 300 K this factor is $\sim 10^{10}$. Due to the more limited sampling of these high-energy conformations, estimates of properties related to these conformations (e.g., transition states) are expected to have associated comparatively large errors (see below, Figure 3D).

In Figure 3B, probability distributions of the remaining dihedral angles $\chi 2$, ϕ, and ψ in the energy sampling run are compared to those in the 1000 K canonical ensemble. As for the dihedral angle $\chi 1$, it is found that low-energy conformations are more likely to be sampled than high-energy conformations. The distributions in the 1000 K canonical ensemble and in the energy sampling run are similar but not equal. This is different from the situation found for the potential energies for which the sampling probabilities in the 1000 K canonical ensemble and in the energy sampling run are significantly different (Figure 2). In both cases the sampling probabilities of the energy sampling run are, however, an average of the sampling probabilities of the canonical ensembles at temperatures between 280 and 2000 K. The difference is that for dihedral angles the sampling probabilities in the 1000 K canonical ensemble, too, are approximately the average of the sampling probabilities of the canonical ensembles at temperatures between 280 and 2000 K; for the potential energy this is not the case.

In Figure 3C, examples are given for estimates of properties not directly related to the potential energy. Shown are estimates of the potentials of mean force for the dihedral angles $\chi 1$, $\chi 2$, ϕ, and ψ. The estimates were derived from the run of Figure 2. To evaluate the accuracy, the estimate for $\chi 1$ is compared, in Figure 3D, to an estimate of the potential of mean force obtained from only the first nanosecond of the same energy sampling run, to an estimate obtained from a different, 1 ns, energy sampling run, and to an estimate obtained using conventional adaptive umbrella sampling of the dihedral angle $\chi 1$.[6] The estimate from conventional adaptive umbrella sampling has been shown to be reproducible and to agree well with estimates obtained using internal coordinate constraints and thermodynamic perturbation theory.[6] The different estimates in Figure 3D agree well in the low-energy part of the potential. In the high-energy part, the results from the short energy sampling runs and, to a lesser degree, the one from the long energy sampling run deviate from the result of conventional adaptive umbrella sampling. This is to be expected since these high-energy parts are sampled less frequently than the low energy parts (see above). Since these high-energy conformations are unimportant at 300 K (Figure 3A) the errors have a negligible effect on properties calculated for the system at 300 K. However, if one were concerned with dynamics, better estimates would be required for the properties of the transition states; these could be obtained with longer simulations or by alternative weighting schemes.

The time-dependence of the sampling is illustrated in Figure 4. After a short initial phase of about 100 ps during which the umbrella potential converges to within 0.5 kcal/mol of the final potential, the potential energy of the system varies rapidly within the allowed range. The system samples a variety of conformations with root-mean-square deviation (RMSD) values relative to the selected low-energy structure (see caption of Figure 4) of up to 0.3 Å for the three backbone heavy atoms and up to 2.4 Å for all atoms (including hydrogens), with a range of about 1.2 Å for the RMSD of the latter. Counting it as one visit if the system returns to within 0.3 Å all-atom RMSD of the selected low-energy structure after it has been in regions of conformational space with an RMSD of more than 1 Å, the system visits the region of conformational space close to the selected low-energy structure 112 times. Since the fluctuations in the potential energy are very rapid (on the picosecond time scale), the number of visits calculated in this way depends on how often conformers are sampled from the trajectory. The

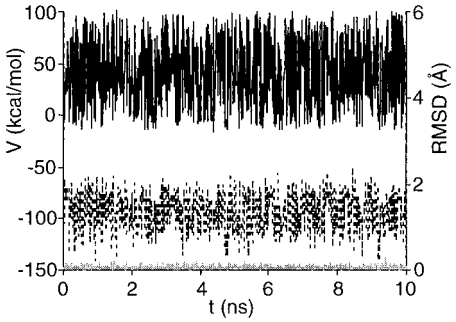

Figure 4. Time evolution of the potential energy and the RMSD from a selected reference conformer in the 10 ns potential energy adaptive umbrella sampling run of the threonine dipeptide; continuous line, potential energy; dashed line, all-atom RMSD; gray line, backbone RMSD. From all the conformers found in the run, the one with the lowest energy was selected as reference. Values are plotted for conformers at 10 ps intervals.

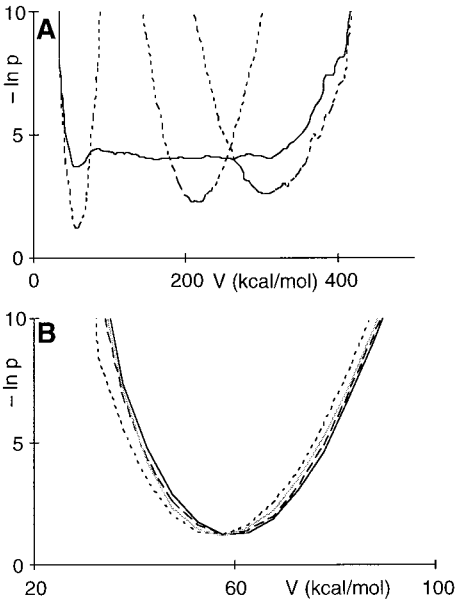

Figure 5. Probability distributions of the potential energy of the met-Enkephalin system. (A) Probability distribution in the second half of the potential energy adaptive umbrella sampling run at 300 K (continuous line), and distribution in the canonical ensembles at 300, 1000, and 1500 K (dashed lines). (B) Probability distributions of the potential energy in the canonical ensemble at 300 K calculated from the data of the different energy sampling runs; black and gray continuous line, the two 2 ns runs at 1000 K; long and short dashes, first and second half of the 4 ns run at 300 K.

number of 112 visits was obtained for conformers taken every 0.1 ps (In Figure 4, RMSD values are plotted for conformers taken every 10 ps).

4.2. Met-Enkephalin. The probability distributions of the potential energy of met-Enkephalin in one of the energy sampling runs and in the canonical ensembles at 300, 1000, and 1500 K are shown in Figure 5A. In the energy sampling run, potential energies are sampled approximately uniformly. The modification of the umbrella potential according to eq 28 restricts sampling to the potential energies occurring in canonical ensembles at a temperature between T_{min} (280 K) and T_{max} (1500 K). The modifications of the umbrella potential according to eq 25 is irrelevant in this case since potential energies larger than $V_{max} = 500$ kcal/mol or smaller than $V_{min} = -500$ kcal/mol do not occur. Some small systematic deviations from the ideal flat distribution are caused by the limited number of functions used to represent the continuous umbrella potential. In general, the deviations between the discrete and the fitted continuous potential are less than 0.03 kcal/mol. At potential energies around 57 kcal/mol, the fitted potential is systematically too small by about 0.13 kcal/mol; at potential energies around 77 kcal/mol it is too large by about 0.06 kcal/mol. The difference between the discrete and the fitted continuous potential has no effect on the properties calculated from the run since the WHAM eqs 1−4 are valid for arbitrary umbrella potentials as long as the same umbrella potential is used in the molecular dynamics simulation and in eq 3. The most likely potential energies in the canonical ensemble at 300, 1000, and 1500 K are ∼ 57.5, ∼217.5, and ∼302.5 kcal/mol, respectively. The potential energy of the conformer with the lowest energy sampled in the three energy sampling runs is 33.6 kcal/mol; this is referred to as the "selected low-energy structure" and used as the reference structure in subsequent analysis. The selected low-energy structure is not the structure with the lowest possible potential energy, and many other structures exist with a similar potential energy (see below and Figure 6), including some with a different conformation. Energy minimization of this selected low-energy structure lowered the energy to −8.9 kcal/mol. The dihedral angles of this energy-minimized conformer are listed in Table 2.

To assess the reproducibility of the probability distributions, we recalculated the distributions for the canonical ensemble at 300 K from the two 2 ns runs at 1000 K and from the first and second half of the 4 ns run at 300 K. These distributions are compared in Figure 5B. Agreement is good but not perfect. For example, the probability for potential energies between 45 and 50 kcal/mol is estimated to be 5%, 9%, 7%, or 12%, respectively.

Dynamic aspects of the sampling are illustrated in Figure 6; A and B are from the two 2 ns runs at 1000 K, C is from the first 2 ns, and D from the last 2 ns of the 4 ns run at 300 K. Considering all three simulations together, the system visits the region of conformational space close to the selected low-energy structure six times (the backbone RMSD is smaller than 0.5 Å at 1002−1080 ps and 1985−2000 ps in Figure 6A, at 1830−1897 ps and 1928−1934 ps in Figure 6B, at 1309−1310 ps in Figure 6C, and at 3276−3355 ps in Figure 6D). The individual simulations are too short to ensure that the system revisits the same region of conformational space several times. Since the precision depends on how often conformations with a large contribution to the estimate are sampled, some results derived from a single 2 ns may still have significant errors and longer runs would be required to obtain accurate values for all equilibrium properties of the system.

The continuous line in Figure 7A shows the density of states $n(V)$ (eq 21). The dependence of the logarithm of the density of states is well approximated with a quadratic function (dashed line in Figure 7A). It is of interest that such a small system obeys this relationship, which can be derived from the random energy model[47] that has been applied to polymers and to protein models.[47] An analogous result has been obtained for the distribution of the low-energy conformers of cycloheptadecane.[48]

Figure 7B illustrates the evolution of the adaptive umbrella potential during each of the runs. The umbrella potential depends on the simulation temperature which was either 300 or 1000 K (see above). The density of states $n(V)$ (Figure 7A) is independent of the simulation temperature and can be calculated from the umbrella potential $U_i(V)$ (eq 21). We therefore plot the logarithm of the ratio of the density of states at $V_1 = 57.5$ kcal/mol and at $V_2 = 217.5$ kcal/mol, the most

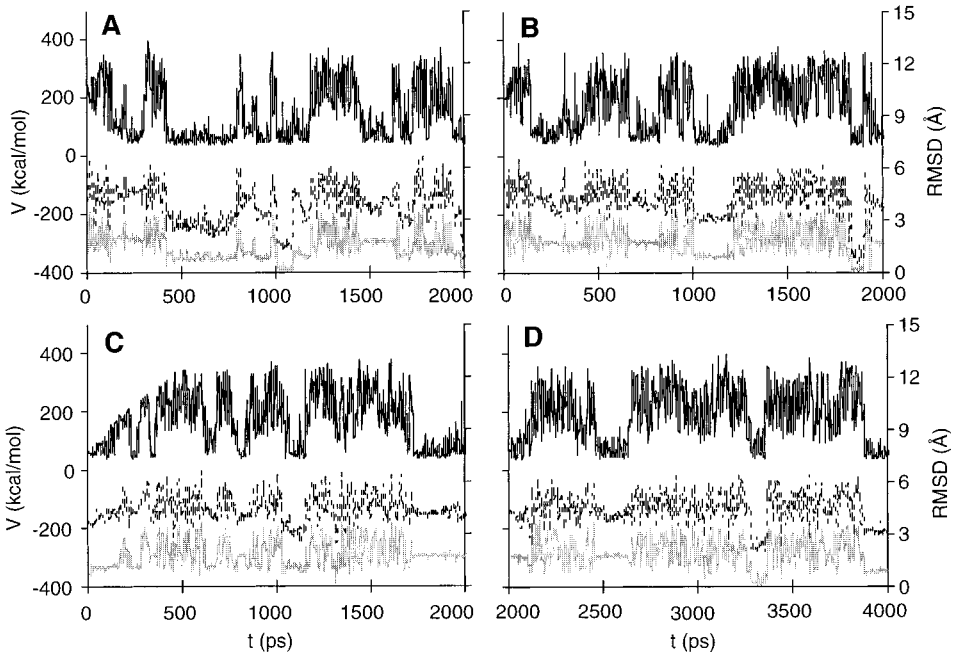

Figure 6. Time evolution of the three potential energy adaptive umbrella sampling runs of met-Enkephalin. The potential energy and the RMSD from the selected low-energy conformer are shown (see text); continuous black line, potential energy; dashed line, all-atom RMSD; gray line, backbone RMSD. A and B are the two 2 ns runs at 1000 K; C and D are the first and second half of the 4 ns run at 300 K.

likely potential energies in canonical ensemble at 300 K and at 1000 K, respectively (see above),

$$\alpha = \ln\left(\frac{n(V_2)}{n(V_1)}\right) = \frac{1}{RT}(V_2 - V_1) + \frac{1}{RT}(U_i(V_2) - U_i(V_1)) \quad (36)$$

where T, the simulation temperature of the energy sampling run, is either 300 or 1000 K, respectively, and eq 21 was used to obtain the right-hand side of eq 36. After 200 ps (10 updates of the umbrella potential) the values of α in the three runs differ by less than 12, after 2 ns by less than 2. During the last 2 ns of the 4 ns run at 300 K, α varies by less than 0.8, which means (eq 36) that the umbrella potential varies by less than $0.8RT$. Thus, 2 ns of simulation seem to be sufficient for the umbrella potential to converge, independently of whether the simulations are carried out at 300 K or at 1000 K.

In Figure 8A, the contributions of the different functions used to represent the continuous umbrella potential are analyzed, and in Figure 8B, the convergence of the contributions is illustrated for some of the functions. The data are from the 2 ns run at 1000 K shown in Figure 6A. It is seen that the dominant contributions are from the three polynomial functions (eq 35 with $n = 0, 1, 2$) and from the first few trigonometric functions (eq 34 with small n). The contributions of trigonometric functions with $n > 30$ are more than 1000 times smaller than the largest contribution. The convergence of the contributions during the adaptive umbrella sampling run is fast for functions with important contributions to the final potential, e.g., the contribution of the sine term (eq 34) with $n = 1$ is at 94% of its final value after 25 updates. The contributions of the less important functions converge more slowly. Although the absolute fluctuation of these contributions are small (less than 1% of the largest contribution), the fluctuations relative to the individual final values, which are plotted in Figure 8B, are large. This instability in the convergence suggests that a smaller number of functions in the expansion could be adequate as well.

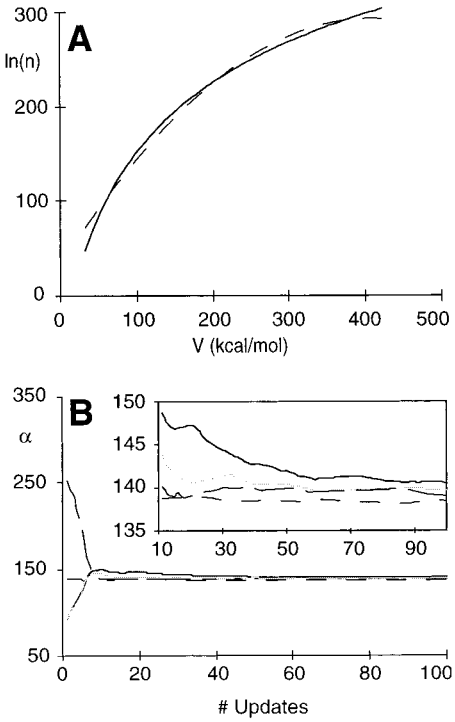

Figure 7. (A) Logarithm of the density of states (eq 21) calculated from the second half of the 4 ns energy sampling run at 300 K. The continuous line shows the data obtained with eq 21; the dashed line is the quadratic function $31.29 + 1.28V - 0.0156V^2$ which was obtained by a least-squares fit (correlation coefficient 0.9956) to the data of the continuous line. (B) Evolution of the umbrella potential during the three potential energy adaptive umbrella sampling runs of met-Enkephalin; black and gray continuous line, runs at 1000 K; long and short dashes, first and second half of the run at 300 K. Plotted is the value α (eq 36) versus the number of updates of the umbrella potential. The inset shows an expanded version of the convergence behavior after the first 10 updates.

From the trajectories of the three runs, various average properties of met-Enkephalin were calculated. In Figure 9, distributions of the backbone dihedral angles ϕ and ψ are

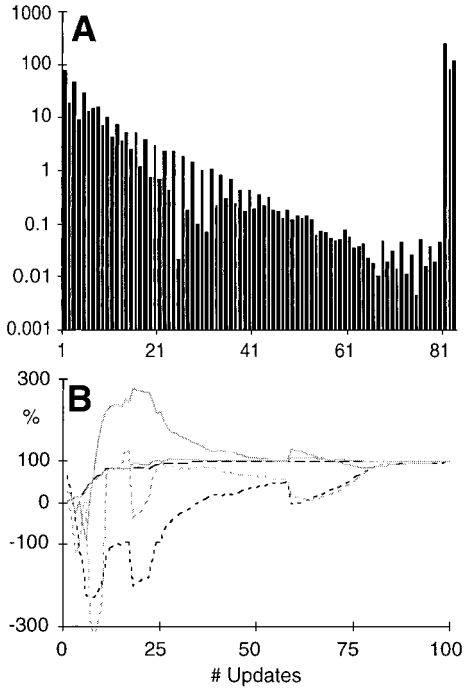

Figure 8. (A) Contribution of the different fit functions (eqs 34 and 35) to the continuous umbrella potential. The first 40 pairs of bars show the contributions of the sine and cosine functions (eq 34, small values of n are plotted first). The last three bars are for the contributions of the polynomial functions (eq 35, small values of n are plotted first). (B) Convergence of the contributions of different functions to the continuous umbrella potential. Plotted is the contribution relative to the final value at the end of the run versus the number of updates of the umbrella potential. Continuous black line, black line with short dashes, continuous gray line, and gray line with short dashes, sine functions (eq 34) with $n = 1$, 11, 21, 31, respectively; black line with long dashes, polynomial function with $n = 3$ (eq 35).

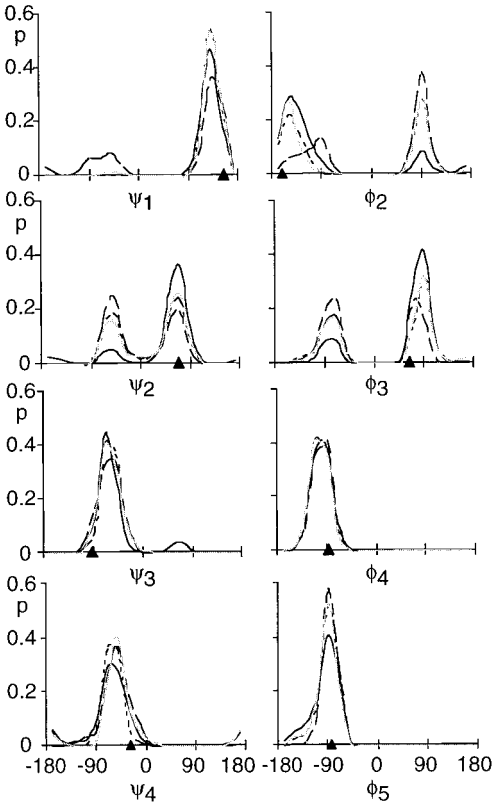

Figure 9. Distribution of the backbone dihedral angles of met-Enkephalin at 300 K calculated from the three potential energy adaptive umbrella sampling runs of met-Enkephalin; black and gray continuous line, runs at 1000 K; long and short dashes, first and second half of the run at 300 K. Triangles mark the angles of the minimum energy conformer from ref 23 listed in Table 2. This conformer was obtained using multicanonical Monte Carlo simulations with the ECEPP/2[49,50] force field with fixed bond lengths and angles.

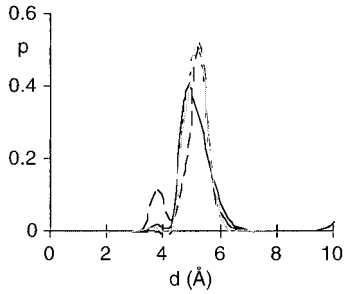

Figure 10. Distribution at 300 K of the distance between the amide nitrogen of Tyr-1 and the carboxyl carbon of Met-5 of met-Enkephalin calculated from the three potential energy adaptive umbrella sampling runs; black and gray continuous line, runs at 1000 K; long and short dashed, first and second half of the run at 300 K.

presented (for the definition see Figure 1B), and in Figure 10, we show the distribution of the distance between the amino and carboxy terminus. The distributions from the different runs agree in that very similar conformations are predicted to be important. However, the estimates from different runs show some quantitative differences. To obtain better estimates for observables that depend on sampling different minima, longer simulations with more transitions between the minima are required. Properties that are similar for all relevant minima, e.g., the ϕ angle of Phe-4, are expected to be most precise, since the number of transitions between the minima should have little influence. From the distributions, it can be seen that the dihedral angles of the N-terminal residues Tyr and Gly adopt several distinct conformations, while the remaining dihedral angles exist predominantly in a single conformation. The two termini, even though they are not charged, are close together in most of the conformations. The maxima of the calculated distributions agree well with values of the main chain dihedral angles in the minimum energy conformation (Table 2 and triangles in Figure 9) determined[23] using the ECEPP/2[49,50] force field with fixed bond lengths and angles. The side chain conformations are also similar except for Met-5.

To describe the essential conformational features of met-Enkephalin, it is useful to represent it as a small set of conformers that have significant populations. Such a set of conformers provides information on the correlation among the structural properties of the system that is not available from one-dimensional distributions; e.g., it is possible to identify different patterns of hydrogen bonds (see below). Often, the conformer with the lowest possible potential energy is assumed to give a reasonable description of the system at room temperature. However, for flexible systems, such as the present one, a single conformer does not represent the distribution of important conformations. Adaptive umbrella sampling of the potential energies provides directly the conformations which are important at 300 K. In Figure 11, three representative structures of the run from Figure 6B are shown; their dihedral angles are listed in Table 2. To select these structures, we analyzed the structures sampled every picosecond from the run of Figure 6B and chose the 82 structures with an energy between 55 and 60 kcal/mol; these are the structures which make the dominant contributions to the average properties at 300 K (Figure 5). Hydrogen bonds present in these structures are listed

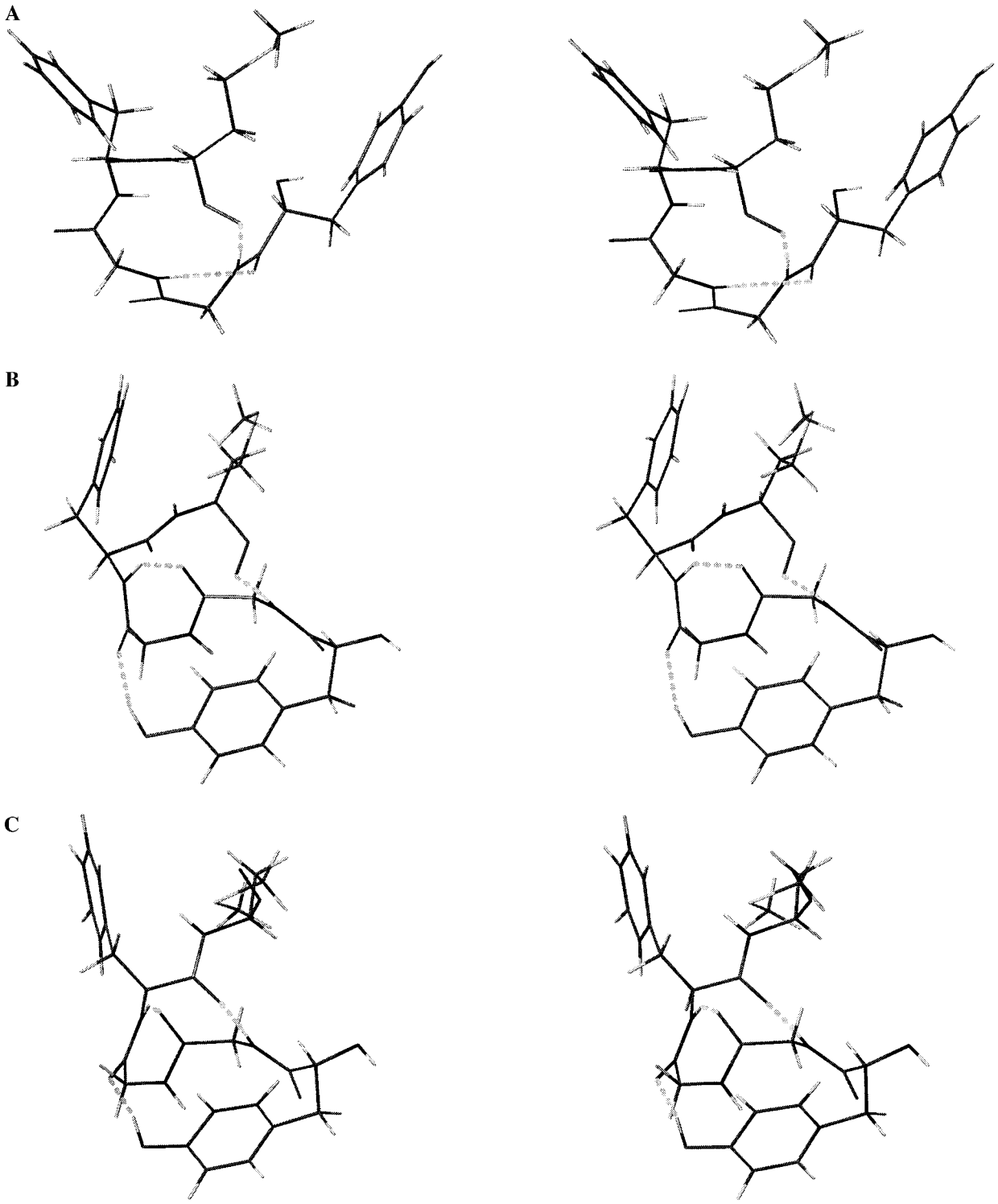

Figure 11. Stereoviews (cross eyed) of three representative conformers selected from the run shown in Figure 6B (see text). Hydrogen bonds are shown with dotted lines. (A) conformer with hydrogen bonds between HN Gly-2 and O Met-5, and between HN Gly-3 and O Tyr-1; (B) conformer with hydrogen bonds between HN Gly-2 and O Met-5, between HN Phe-4 and O Gly-2, and between Hη Tyr-1 and O Gly-3; (C) conformer with hydrogen bonds between HN Gly-2 and O Gly-4, between HN Phe-4 and O Gly-2, and between Hη Tyr-1 and O Gly-3. The figure was prepared with the program MOLMOL.[51]

in Table 3. The 82 structures were further characterized by identifying recurring patterns of hydrogen bonds. Three major patterns (listed in the caption of Figure 11) were found and, based on that, the three representative structures, A, B and C, shown in Figure 11, were selected. Backbone and all-atom RMSD values between pairs of the representative structures are listed in Table 4; they are larger than 0.95 and 2.48 Å, respectively. From the RMSD values (Table 4) and the dihedral angles (Table 2) it is seen that structure C is very similar to the selected low-energy structure.

To assess how well these structures represent the ensemble at 300 K, the distributions of the RMSD from the three representative structures are plotted in Figure 12. These distributions were calculated using eq 17 from the 10^5 structures taken every 20 fs rather than from the selected 82 structures or from the 2000 structures taken every ps. In this large set of

TABLE 3: Hydrogen Bonds Identified in the 82 Low-Energy Met-Enkephalin Structures[a]

donor		acceptor		number[b]
Tyr−1	HN	Met−5	O	2
Tyr−1	ηOH	Gly−3	O	28
Gly−2	HN	Phe−4	O	9
Gly−2	HN	Met−5	O	64
Gly−3	HN	Tyr−1	O	22
Phe−4	HN	Gly−2	O	18
Met−5	HN	Gly−2	O	2

[a] Structures with potential energies between 55 and 60 kcal/mol selected from conformations taken at 1 ps intervals from the energy sampling run shown in Figure 6B. [b] Number of structures with a given hydrogen bond identified according to the criteria $d \leq 2.4$ Å and $\gamma \leq 35°$, where d is the proton-acceptor distance and γ the angle between the donor-proton bond and the line connecting the acceptor and donor heavy atoms. In empirical force fields the bond angle θ between the lines connecting the proton to the acceptor and donor heavy atoms is used to characterize hydrogen bonds. For $d = 2$ Å and a length of 1 Å for the donor-proton bond, $\gamma \leq 35°$ implies $\theta \leq 52°$.

TABLE 4: RMSD between the Three Representative Structures in Figure 11 and the Selected Low-Energy Conformer

RMSD (Å)[a]	structure A	structure B	structure C	low energy[b]
structure A	-	1.94	2.06	1.82
structure B	3.85	-	0.95	1.06
structure C	4.01	2.48	-	0.60
low energy	3.81	2.64	1.34	-

[a] Backbone RMSD values are given above, all-atom RMSD values below the diagonal. [b] Selected low-energy conformer (see text) after energy minimization.

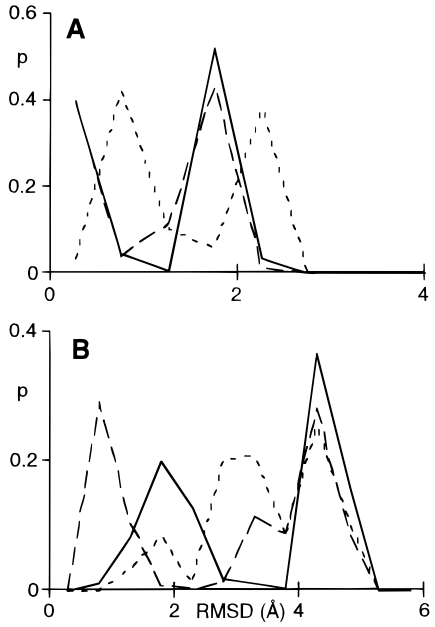

Figure 12. Distributions of structures at 300 K calculated from the run shown in Figure 6B. (A) Backbone RMSD with respect to the structures A−C in Figure 11, (B) corresponding all-atom RMSD; continuous line, long dashes, and short dashes are for structures A, B, and C, respectively.

structures, the normalized weights of individual structures, including the three selected reference structures, are small; the largest normalized weight is 7×10^{-4} and structures A, B, and C have normalized weights 4×10^{-5}, 2×10^{-5}, and 8×10^{-5}, respectively. Figure 12 shows that many of the conformers at 300 K are similar to either structure A or B, and a few are similar to structure C. The values plotted in Figure 12 should be treated with care since no error estimates have been obtained and since it has been found that a single 2 ns simulation is too short to ensure that the important regions of conformational space are visited several times (Figure 6). Comparing the distributions of the RMSD values for the backbone atoms to those for all atoms (Figure 12B), it can be seen that the side chains of conformers similar to structure B are better defined than those of conformers similar to structure A.

TABLE 5: Percentage of Conformations with Backbone RMSD Values Smaller than 0.5 Å from the Three Representative Structures in Figure 11 and from the Selected Low-Energy Conformer

	run 6A[a]	run 6B[a]	run 6C[a]	run 6D[a]	avg[b]	std dev[b]	runs 6A−D[c]
structure A	12	39	54	42	37	18	33
structure B	59	40	15	38	38	18	41
structure C	5	4	0	4	3	2	3
low energy	12	13	0	14	10	6	10

[a] In the columns run 6A, 6B, 6C, and 6D, values are listed which were calculated from the two energy sampling runs at 1000 K and the two halves of the energy sampling run at 300 K of Figure 6, A−D. [b] Average and standard deviation of the values from runs 6A to 6D. [c] Values derived by combining the data of runs 6A to 6D and solving the WHAM eqs 1, 2, 4, and 9 for the combined data.

Improved estimates together with estimates for the errors can be obtained by comparing the results from the different runs. In Table 5, the percentages of conformers are listed that have backbone RMSD values smaller than 0.5 Å with respect to the three representative structures; values derived from the different runs as well as the averages and the standard deviations are given. Dividing the standard deviations by the square root of the number of independent measurements, i.e., by 2, gives an estimate for the errors of the averages. It is found that 37 ± 9%, 38 ± 9%, and 3 ± 1% of the conformers at 300 K have an RMSD value smaller than 0.5 Å with respect to the three structures A−C, respectively. The pairwise backbone RMSD between the structures A−C are larger than 0.9 Å (Table 4). This shows that structures A and B each represent about 40% of the conformers at 300 K, and structure C another few percent. Plots of the potential energy and the RMSD with respect to the three structures versus time (not shown) show that almost all the remaining low-energy conformers have backbone RMSD values smaller than 1 Å with respect to at least one of the three structures. The high-energy conformers, in general, have backbone RMSD values larger than 1 Å with respect to all three structures. Some of these represent transition states for going from one representative structure to another.

Instead of averaging estimates obtained independently from each of the runs where the exact weights given to each run are not clear, it is possible to combine the runs by treating them as a single run. In the former case, the WHAM eqs 1−4 are solved independently and weighting factors (eq 14) are calculated independently for each of the runs. In the latter case, the WHAM eqs are solved for all runs together. This results in more precise values for $\tilde{p}^o_k$ (eq 1) which in turn leads to better estimates for the weighting factors (eq 14). Therefore, solving the WHAM eqs jointly for all runs should give more precise estimates for observables (eq 15); however, there is no easy way to obtain error estimates. Since in the present example, different simulation temperatures were used for the different runs, eq 9 had to be used instead of eq 3 in the WHAM equations. The values obtained in this way are listed in Table 5. They are similar to the averages of the estimates obtained independently for each of the runs.

5. Discussion

We have developed an adaptive umbrella sampling method for molecular dynamics simulations that uses the potential energy as the umbrella parameter. In contrast to most other methods, the present approach does not require a choice of specific degrees of freedom related to the geometry of the system for the umbrella potential. Equilibrium values and distributions of any observable of the system that can be expressed as a function of the Cartesian coordinates can be derived from the simulations. The precision of the derived properties depends on the system and on the property itself. Since the method leads to uniform sampling of the potential energy, properties related directly to the potential energy can be derived with high precision. Because the rate of crossing barriers separating different local minima is enhanced, properties which require sampling of different local minima separated by large free energy barriers can be estimated with higher precision than in an unbiased simulation. Further, sampling of high-energy conformers is enhanced, and properties related to high-energy conformers can be obtained. Application of the method to the dihedral angle distribution function of the threonine dipeptide and met-Enkephalin demonstrates its utility. Although conformers in a given potential energy range have the same probability at high and low potential energies in a converged simulation, part of the equal weighting is due to the much larger density of states associated with the former relative to the latter. This means that individual high-energy conformers are still less likely to be sampled than individual low-energy conformers, and estimates of properties of specific high-energy conformers will in general be less precise than estimates of properties related to low-energy conformers. Thus, the transition state free energies are expected to be less accurate than those of the most important conformers.

From the present results some limitations of the method are evident. For met-Enkephalin, different states can be distinguished in which the system has relatively well-defined properties, i.e., the average and the fluctuations of the potential energy and of the RMSD from the selected low-energy structure. Transitions between the different states are faster than in unbiased simulations but do not occur very often. Thus, the system spends most of the simulation time in different states which are separated by free energy barriers, even after the umbrella potential has converged. For the threonine dipeptide, different states separated by barriers also can be identified in the adaptive umbrella sampling runs (Figure 3, A and B). For example, the $\chi 1$ values $-60°$ and $60°$ are separated by a free energy barrier (Figure 3A). The height of the residual free energy barriers that remain after the umbrella potential has converged determines how often transitions betwen different important states do occur. In the adaptive umbrella sampling runs of met-Enkephalin and the threonine dipeptide, several transitions occur per nanosecond, indicating that the residual free energy barriers are small. For other systems, e.g., for protein folding, residual free energy barriers might be higher. In such cases variables, whose values differ for the significant local minima, the transition state and the unimportant conformations (e.g., selected terms in the potential energy) may be useful for determining the umbrella potential. If appropriately chosen, the variables would lead to sampling of both transition states and local minima conformations without generating too many unimportant conformations. Such extensions of the method are under investigation.

Acknowledgment. The Institut de Développment et des Ressources en Informatique Scientifique (I.D.R.I.S.) provided an allocation of CRAY T3D/T3E computer time which was used for part of this work. C.B. was supported by the EMBO long term fellowship ALTF 448-1995. The laboratory is supported in part by the CNRS (URA 422), by the Ministère de l'Education Nationale, by a grant from the Association pour la Recherche contre le Cancer (France). The work done in the U.S.A. is supported by a grant from the National Institutes of Health (U.S.A.).

References and Notes

(1) Brooks, C. L., III; Karplus, M.; Pettitt, B. M. *Proteins: A Theoretical Perspective of Dynamics, Structure, and Thermodynamics*. John Wiley & Sons: New York, 1988.
(2) McCammon, J. A.; Harvey, S. *Dynamics of Proteins and Nucleic Acids*; Cambridge University Press: Cambridge, U.K., 1987.
(3) Elber, R. *Curr. Opin. Struct. Biol.* **1996**, *6*, 232.
(4) Boczko, E. M.; Brooks, C. L., III *Science* **1995**, *269*, 393.
(5) Boczko, E. M.; Brooks, C. L., III *J. Phys. Chem.* **1993**, *97*, 4509.
(6) Bartels, C.; Karplus, M. *J. Comput. Chem.* **1997**, *18*, 1450.
(7) Ferrenberg, A. M. Thesis, Carnegie-MELLON University, Pittsburgh, PA, Order Number 9006207, 1989.
(8) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J. Comput. Chem.* **1992**, *13*, 1011.
(9) Torrie, G. M.; Valleau, J. P. *J. Comput. Phys.* **1977**, *23*, 187.
(10) Hooft, R. W. W. *J. Chem. Phys.* **1992**, *97*, 6690.
(11) Mezei, M. *J. Comput. Phys.* **1987**, *68*, 237.
(12) Straatsma, T. P.; McCammon, J. A. *J. Chem. Phys.* **1994**, *101*, 5032.
(13) Tsujishita, H.; Moriguchi, I.; Hirono, S. *J. Phys. Chem.* **1993**, *97*, 4416.
(14) Liu, Z.; Berne, B. J. *J. Chem. Phys.* **1993**, *99*, 6071.
(15) Roitberg, A.; Elber, R. *J. Chem. Phys.* **1991**, *95*, 9277.
(16) Verkhivker, G.; Elber, R.; Nowak, W. *J. Chem. Phys.* **1992**, *97*, 7838.
(17) Piela, L.; Kostrowicki, J.; Scheraga, H. A. *J. Phys. Chem.* **1989**, *93*, 3339.
(18) Straub, J. E.; Ma, J.; Amara, P. *J. Chem. Phys.* **1995**, *103*, 1574.
(19) Lee, J. *Phys. Rev. Lett.* **1993**, *71*, 211.
(20) Lyubartsev, A. P.; Martsinovski, A. A.; Shevkunov, S. V.; Vorontsov-Velyaminov, P. N. *J. Chem. Phys.* **1992**, *96*, 1776.
(21) Berg, B. A. *Int. J. Mod. Phys.* **1992**, *C3*, 1083.
(22) Berg, B. A.; Neuhaus, T. *Phys. Rev. Lett.* **1992**, *68*, 9.
(23) Hansmann, U. H. E.; Okamoto, Y. *J. Comput. Chem.* **1993,** *14*, 1333.
(24) Hao, M.-H.; Scheraga, H. A. *J. Phys. Chem.* **1994**, *98*, 4940.
(25) Kidera, A. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 9886.
(26) Hansmann, U. H. E.; Okamoto, Y. *Physica A* **1994**, *212*, 415.
(27) Bradbury, A. F.; Smyth, D. G.; Snell, C. R. *Nature* **1976**, *260*, 165.
(28) Camerman, A.; Mastropaolo, D.; Karle, I.; Karle, J.; Camerman, N. *Nature* **1983**, *306*, 447.
(29) Nakajima, N.; Nakamura, H.; Kidera, A. *J. Phys. Chem.* **1997**, *101*, 817.
(30) Bartels, C.; Stote, R. H.; Karplus, M. Manuscript in preparation.
(31) Kumar, S.; Payne, P. W.; Vásquez, M. *J. Comput. Chem.* **1996**, *17*, 1269.
(32) Gelman, A.; Carlin, J. B.; Stern, H. S.; Rubin, D. B. *Bayesian Data Analysis*; Chapman & Hall: London, 1995.
(33) Krengel, U. *Einführung in die Wahrscheinlichkeitstheorie und Statistik*. Vieweg: Braunschweig/Wiesbaden, Germany, 1991.
(34) Atkins, P. W. *Physical Chemistry*; Oxford University Press: Oxford, 1987.
(35) Huang, K. *Statistical Mechanics*. John Wiley & Sons: New York, 1987.
(36) Karplus, M.; Shakhnovich, E. In *Protein Folding: Theoretical Studies of Thermodynamics and Dynamics*; Creighton, T. E., Ed.; W. H. Freeman and Company: New York, 1992.
(37) Sali, A.; Shakhnovich, E.; Karplus, M. *Nature* **1994**, *369*, 248.
(38) Valleau, J. P.; Torrie, G. M. In *A Guide to Monte Carlo for Statistical Mechanics: 2. Byways*; Berne, B. J., Ed.; Plenum Press: New York, 1977.
(39) Berg, B. A.; Hansmann, U. H. E.; Okamoto, Y. *J. Phys. Chem.* **1995**, *99*, 2236.
(40) Hansmann, U. H. E.; Okamoto, Y.; Eisenmenger, F. *Chem. Phys. Lett.* **1996**, *259*, 321.
(41) Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695.
(42) Nosé, S. *J. Chem. Phys.* **1984**, *81*, 511.

(43) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.
(44) MacKerell, A. D., Jr.; et al. *FASEB J.* **1992**, *6*, A143.
(45) MacKerell, A. D., Jr.; at al. Manuscript in preparation.
(46) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327.
(47) Shakhnovich, E. I.; Gutin, A. M. *Nature* **1990**, *346*, 773.
(48) Ngo, J. T.; Karplus, M. *J. Am. Chem. Soc.* **1997**, *119*, 5657.
(49) Némethy, G.; Pottle, M. S.; Scheraga, H. A. *J. Phys. Chem.* **1983**, *87*, 1883.
(50) Sipple, M. J.; Némethy, G.; Scheraga, H. A. *J. Phys. Chem.* **1984**, *88*, 6231.
(51) Koradi, R.; Billeter, M.; Wüthrich, K. *J. Mol. Graphics* **1996**, *14*, 51.