

Article

## Transition Pathway and Its Free-Energy Profile: A Protocol for Protein Folding Simulations

In-Ho Lee <sup>1</sup>, Seung-Yeon Kim <sup>2</sup> and Jooyoung Lee <sup>3,\*</sup>

<sup>1</sup> Korea Research Institute of Standards and Science, Daejeon 305-340, Korea;  
E-Mail: ihlee@kriss.re.kr

<sup>2</sup> School of Liberal Arts and Sciences, Korea National University of Transportation,  
Chungju 380-702, Korea; E-Mail: sykimm@ut.ac.kr

<sup>3</sup> Center for In Silico Protein Science, School of Computational Sciences,  
Korea Institute for Advanced Study, Seoul 130-722, Korea

\* Author to whom correspondence should be addressed; E-Mail: jlee@kias.re.kr;  
Tel.: +82-2-958-3869; Fax: +82-2-958-3890.

Received: 15 April 2013; in revised form: 22 July 2013 / Accepted: 29 July 2013 /

Published: 2 August 2013

---

**Abstract:** We propose a protocol that provides a systematic definition of reaction coordinate and related free-energy profile as the function of temperature for the protein-folding simulation. First, using action-derived molecular dynamics (ADMD), we investigate the dynamic folding pathway model of a protein between a fixed extended conformation and a compact conformation. We choose the pathway model to be the reaction coordinate, and the folding and unfolding processes are characterized by the ADMD step index, in contrast to the common *a priori* reaction coordinate as used in conventional studies. Second, we calculate free-energy profile as the function of temperature, by employing the replica-exchange molecular dynamics (REMD) method. The current method provides efficient exploration of conformational space and proper characterization of protein folding/unfolding dynamics from/to an arbitrary extended conformation. We demonstrate that combination of the two simulation methods, ADMD and REMD, provides understanding on molecular conformational changes in proteins. The protocol is tested on a small protein, penta-peptide of met-enkephalin. For the neuropeptide met-enkephalin system, folded, extended, and intermediate states are well-defined through the free-energy profile over the reaction coordinate. Results are consistent with those in the literature.

**Keywords:** molecular dynamics; free energy; reaction coordinate

---

## 1. Introduction

Understanding the protein folding process inside a living cell is one of the most important issues in modern biology [1,2], and an accurate description of the protein folding pathway has remained a challenging problem in molecular biology for the last few decades. Obtaining both theoretical and experimental information about folding processes as well as transiently populated intermediate states located between folded and unfolded states, requires the development of new theoretical and experimental methods. In recent years, computational approaches based on the energy function of a protein conformation have provided valuable knowledge about the protein folding mechanism along with various experimental techniques such as kinetics study with circular dichroism measurement, NMR spectroscopy, X-ray crystallography, and/or Fersht  $\Phi$  value analysis [3–5]. The reliability of a folding simulation may depend on the adequacy of the potential energy function used. Therefore, to stay in touch with reality, theoretical calculations should be tested against experimental data as much as possible.

When performing a simulation, although there are issues with the accuracy of potential energy function used, one important factor to overcome is that the simulation time should be long enough compared to the actual reaction time of the phenomenon under investigation. Therefore, developing a computational method that either extends the time scale of the simulation or expedites the reaction itself is rather important. In order to draw reliable and consistent conclusions out of a simulation, the overall results of the simulation, especially on kinetic and thermodynamic properties of a protein, should not depend on the length of the simulation time scale. Thus, through rigorous folding pathway simulations, one can provide meaningful assessment of the relevance of a potential energy function used, and if necessary, modification or fine-tuning of the energy function becomes feasible to describe real proteins more accurately.

Recently, various algorithms and hardware for molecular simulations have been developed and applied to protein folding studies. However, carrying out justifiable molecular simulations of proteins is rather demanding, and such simulations are limited to relatively small proteins. One major bottleneck for computational study is the time-scale gap between simulations and experiments. Shaw's group has developed a specialized supercomputer, called Anton, which greatly accelerates the execution of atomistic molecular dynamics (MD) simulations [6]. Recently, Lindorff-Larsen *et al.* [7] used state-of-the-art MD simulations to investigate the folding mechanisms of 12 proteins. They found that during barrier crossing, the long-range contacts that establish the protein's overall fold form early along with the formation of a considerable amount of secondary structures and surface burial. Another difficulty is that, in general, it is difficult to define a good reaction coordinate for description of protein folding/unfolding dynamics from/to an arbitrary extended conformation [8].

In this work, we take two separate algorithms for the investigation of the protein folding mechanism and thermodynamic analysis without much fine-tuning. Recently, we have developed and tested an algorithm that provides a protein folding pathway model with the boundary-value formulation of

Newtonian dynamics [9–13]. The first algorithm we use is the action-derived molecular dynamics (ADMD) method [9,10], and it is used to generate protein folding pathways. Low energy-barrier transition pathways are generated by carrying out ADMD simulations. The advantage of ADMD lies in the fact that it does not require the selection the reaction coordinate in advance. ADMD trajectories are generated by optimizing an action with a specific choice of the target energy. Generated atomic trajectories are shown to be rather close to the Verlet trajectory for the whole trajectory even with a relatively large time step.

The second algorithm we use is the replica-exchange molecular dynamics (REMD) method [14,15], which is used for the thermodynamic analysis of the pathway. Using the pathway obtained by ADMD simulations, we estimate the free-energy profile along the pathway. The conformations obtained from the ADMD simulation provide a set of diverse conformations for the subsequent REMD simulation. Then, we use the REMD method to quantify the thermodynamic properties of protein, *i.e.*, free-energy profile and transition states along the pathway.

In this work, we propose a methodology to construct free-energy profile along the folding pathway of a protein by combining ADMD and REMD methods. The approach is tested to simulate the folding process of a small peptide met-enkephalin. In the following sections, first we introduce ADMD, REMD, and related descriptions on details. Next, we discuss the folding mechanism of met-enkephalin obtained by the combined simulations of ADMD and REMD. The results are compared with existing theories and experiments. A summary is given in the conclusion section.

## 2. Results and Discussion

Before analyzing details of ADMD simulation results, the limitation of ADMD approach should be discussed. The goal of ADMD is to obtain the most probable pathway connecting two end structures. Here the choice of end structures can be problematic, especially since the unfolded state cannot be fully represented by a single structure. Here we used an extended structure as the representative of the unfolded state, but a rigorous treatment should consider multiple unfolded structures. Likewise, obtaining the most probably pathway is not sufficient enough to understand the general folding phenomenon, especially for studying small peptides. However, even with these limitations, proper application of ADMD can provide an insight on the folding mechanism as discussed in this section below.

Here we describe the folding pathway of met-enkephalin generated by combining ADMD with REMD. First, we describe the folding pathway of met-enkephalin investigated by ADMD. Second, we investigate the free-energy profile along the folding pathway generated by ADMD. The free energy profile of a folding process can strongly depend on the choice of the reaction coordinate. By choosing the series of conformational changes obtained by ADMD simulation as the reaction coordinate, we obtain the free-energy profile along the ADMD trajectory.

Met-enkephalin (Tyr-Gly-Gly-Phe-Met) is a small neurotransmitter peptide with morphine-like activity and it is found in the central nervous system. Met-enkephalin has been the object of many computational studies of identifying its low-energy conformations [16–20]. NMR spectroscopy experiments on the peptide in dimethyl sulfoxide support the presence of a turn at Gly<sup>3</sup> and Phe<sup>4</sup> with a hydrogen bond between Gly<sup>2</sup> and Met<sup>5</sup> [21–24]. A recent NMR experiment suggests that the solution structure of met-enkephalin constitutes an ensemble of diverse structures including extended

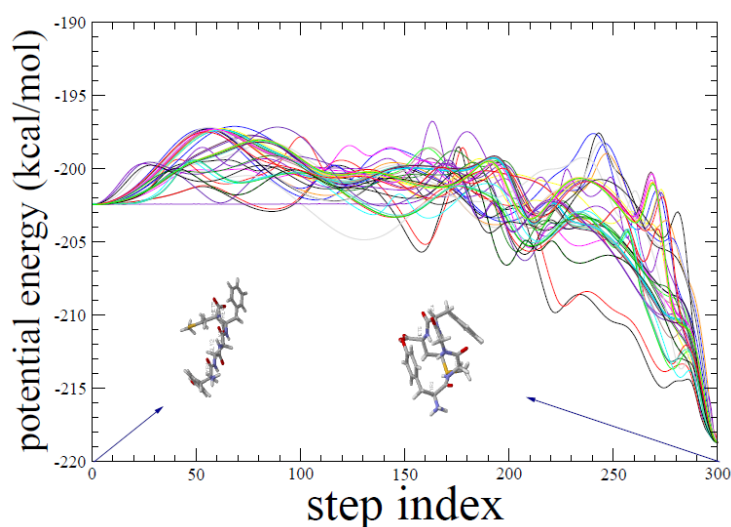
conformations with large fluctuations [25]. No unique native structure has been determined for met-enkephalin in infrared and fluorescence experiments [26,27]. Similar conclusions are obtained by circular dichroism and absorption study of met-enkephalin in solution [28,29]. On the other hand, an extended conformation is obtained by X-ray crystallography [30,31], which is probably due to the crystal packing effect of the peptide.

Identification of the minimum-energy conformation of met-enkephalin has been an issue in the literature, especially in computational studies. However, finite temperature effects on the conformational variation of the small peptide should be properly considered. The flexibility of the peptide is also related to the fact that it binds to at least two receptors of the central nervous system. In addition, met-enkephalin's conformation is medium-dependent. Thus, a study on the multiple stable structures of the peptide and dynamic connectivity between these structures can be a biologically important subject. In our previous study, folding dynamics of met-enkephalin is investigated by ADMD method [32]. We used the same potential and ADMD parameters as in [32]. Here, we have performed additional new ADMD simulations. No data from [32] are used in the analysis of present work. Using only two input structures of the initial (extended) and the final (folded) conformations, dynamic pathways between the two given conformations have been successfully obtained. In the current work, by studying the free-energy profile along the ADMD folding pathway, we find that there exists an ensemble of populated structures along the pathway.

### *2.1. ADMD Simulations: Pathway Analysis and Reaction Coordinate*

We have performed 50 independent ADMD simulations, each of which produced a relatively low potential-energy pathway with an energy barrier in the range of 2.4–6.0 kcal/mol. Initial pathways for ADMD simulations were prepared in a random fashion to sample as many variations as possible in the pathway. In Figure 1, 50 potential-energy fluctuations are shown along the ADMD step index. The same set of initial and final conformations is used for all ADMD simulations. The potential-energy fluctuations clearly show the existence of multiple folding pathways for met-enkephalin. In Figure 2, we show the lowest energy-barrier pathway out of 50. In the formalism of ADMD, low-energy barrier trajectories serve as probable transition pathways, and we assume that the lowest-potential-energy pathway represents the most relevant pathway. This pathway model is used for a further analysis in the present work.

**Figure 1.** Fifty independent potential-energy variations for met-enkephalin are shown along the action-derived molecular dynamics (ADMD) step index. The same initial and final conformations are used for all ADMD simulations. The potential-energy variations clearly show the existence of multiple pathways. There exist many alternative folding routes that connect the extended and compact peptide conformations. Two conformations used as inputs for ADMD simulations are shown. The initial (final) conformation shown at  $j = 0$  (300) is prepared by applying local energy minimization to the fully-extended (compact) structure. We observe that one of the ADMD trajectories is of almost constant potential energy in the first half of the trajectory. In this part, the kinetic energy is rather high. It should be noted that the total energy (sum of kinetic energy and potential energy) is conserved along all ADMD pathways.

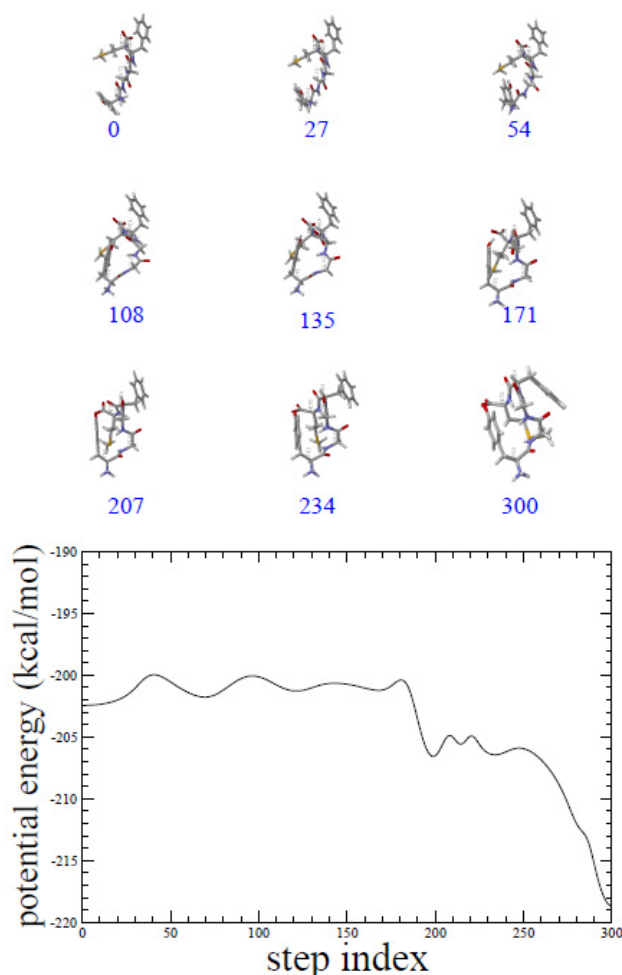


Since atomic motions in the ADMD pathway model statistically satisfy the microscopic time reversal symmetry imposed on the Verlet trajectory, the current ADMD formulation can describe forward as well as reverse reactions (folding and unfolding events). In the current ADMD simulations, calculated trajectories deviate from the Verlet trajectory with minimal error as discussed in the previous section.

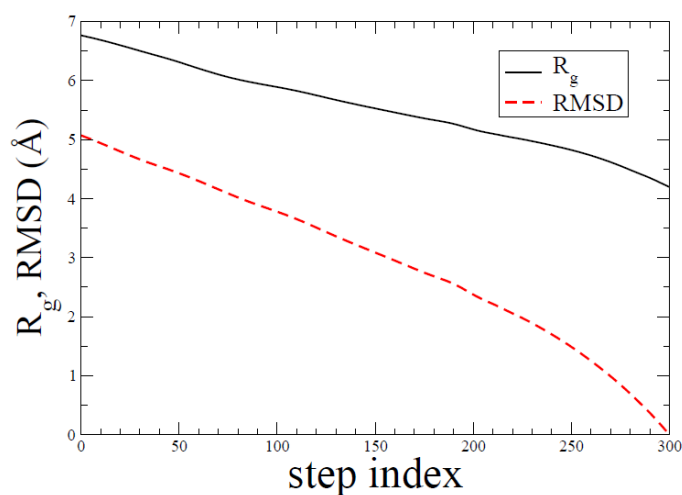
A set of characteristic quantities along the folding pathway is shown in Figure 3. Radius of gyration ( $R_g$ ) and root-mean square deviation from the compact folded conformation are shown. For RMSD calculations, we used all atoms. We have used the quaternion method [33,34] to superimpose two given structures.

A typical distribution of the scalar “error variables” introduced in the previous section is shown in Figure 4. An example for measuring the path quality of a transition path connecting two given configurations is shown. Here, for the pathway generation,  $\Delta = 8.063$  fs,  $N = 75$  atoms, and  $P = 300$  are used. Therefore, a total of  $3N(P - 1) = 67275$  scalar “error variables” are used to obtain the distribution function.

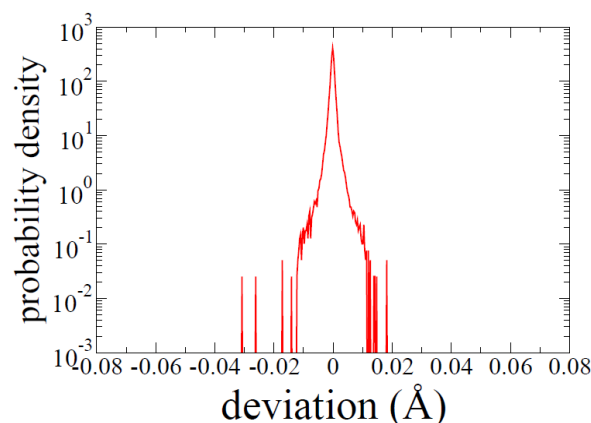
**Figure 2.** Snapshots of the lowest potential-energy-barrier pathway for met-enkephalin are shown. Potential-energy fluctuation is shown along the ADMD step index. As the index increases, the potential-energies related to the pathway model undergo fluctuation. Potential-energy fluctuations associated with the most-fit (lowest potential-energy barrier profile) pathway from the simulations are shown.



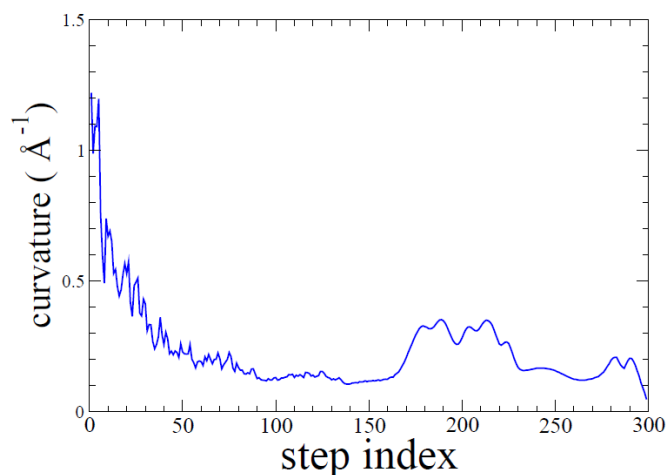
**Figure 3.** Calculated  $R_g$  and root-mean square deviation (RMSD) values in Å are shown along the ADMD step index. We observe that  $R_g$  and RMSD are correlated.



**Figure 4.** A typical distribution of the scalar “error variables” introduced in the text is shown. Here, for pathway generation,  $\Delta = 8.063$  fs,  $N = 75$  atoms, and  $P = 300$  are used. To obtain the distribution function,  $3N(P - 1) = 67275$  scalar “error variables” are used. Calculated distribution of the scalar “error variables”  $\{\bar{\epsilon}_{I,j}\} = 2\bar{q}_{I,j} - \bar{q}_{I,j-1} - \bar{q}_{I,j+1} - \frac{\Delta^2}{m_I} \frac{\partial V(\{\bar{q}_j\})}{\partial \bar{q}_{I,j}}$  is shown. Even with the relatively large time step of  $\Delta = 8.063$  fs, the size of the deviation is rather limited.



**Figure 5.** Curvature of the transition pathway calculated by Equation (5) is shown. The radius of curvature ranges between 1.6 and 10.0 Å, and after the step index  $j = 30$  the radius is larger than 3 Å. This means that there exist no abrupt changes of atomic motions along the path.



Generated ADMD pathways are smooth. That is, no abrupt changes of the conformation are observed. To demonstrate this, we have calculated the curvature of the transition pathway along the ADMD step index (Figure 5). The curvature is defined as follows.

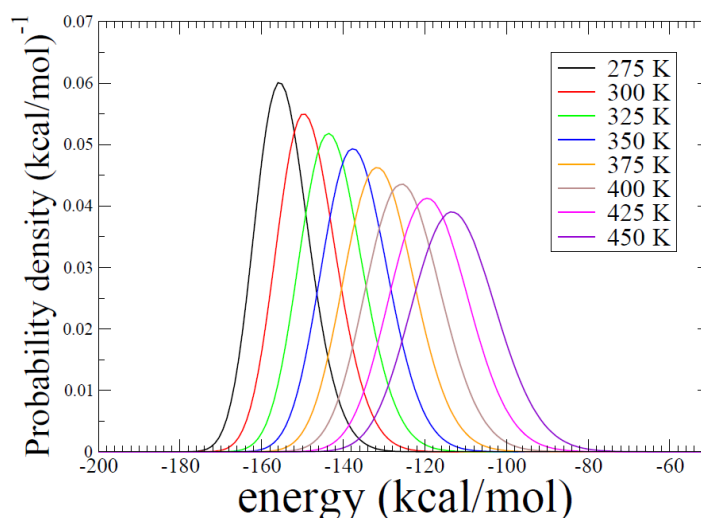
$$C(j) = \frac{\sqrt{\sum_{I=1}^N (2\bar{q}_{I,j} - \bar{q}_{I,j-1} - \bar{q}_{I,j+1}) \cdot (2\bar{q}_{I,j} - \bar{q}_{I,j-1} - \bar{q}_{I,j+1})}}{\sqrt{\sum_{I=1}^N (\bar{q}_{I,j} - \bar{q}_{I,j-1}) \cdot (\bar{q}_{I,j} - \bar{q}_{I,j-1})} \sqrt{\sum_{I=1}^N (\bar{q}_{I,j} - \bar{q}_{I,j+1}) \cdot (\bar{q}_{I,j} - \bar{q}_{I,j+1})}}. \quad (1)$$

The curvature of a straight line is zero, and the curvature of a circle is inversely proportional to its radius. In Figure 5, we find that the curvature of the trajectory stays in the range of  $0.1\text{--}0.6\text{ \AA}^{-1}$ . This means that the radius of curvature associated with the pathway is in the range of  $1.6\text{--}10.0\text{ \AA}$ . In summary, the generated ADMD pathway model, presented as a series of similar conformations, does not undergo any abrupt changes.

## 2.2. REMD Simulations: Thermodynamic Analysis

By using REMD simulations we have obtained probability density functions (PDFs) for a set of temperatures as shown in Figure 6. In the figure, we observe a significant amount of energy overlaps in the coverage of PDF between two adjacent temperatures  $T_i$  and  $T_{i+1}$ .

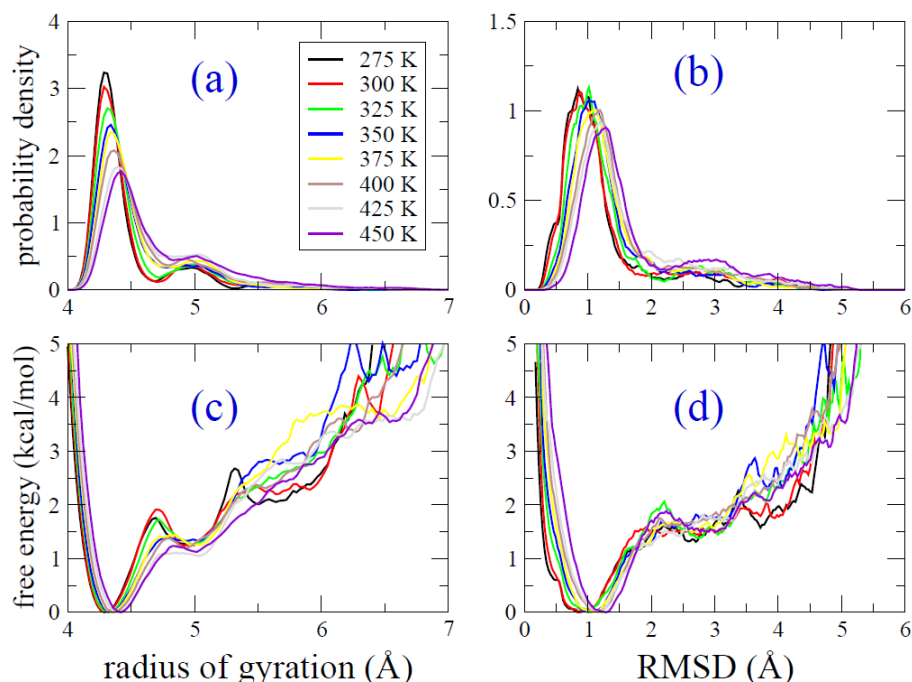
**Figure 6.** Probability density functions obtained from the REMD simulation are shown for eight temperatures. Significant overlaps are observed between neighboring temperatures.



Radius of gyration distribution as a function of temperature obtained from REMD simulation is shown in Figure 7a. The average of radius of gyration increases as temperature increases. Similarly, RMSD distribution as a function of temperature obtained from REMD simulation is also shown in Figure 7b. Again, we find that the average RMSD increases as temperature increases, indicating that the population of the extended-like structures increases. Free energy against the radius of gyration and RMSD are shown in Figure 7c,d, respectively.

In Figure 8, a free-energy profile projected on the 34 reference conformations is shown. A set of free-energy valleys (reference indices, 1, 10, 19, 25 and 34) can be defined along the profile. These valleys represent peptide conformations with accumulated population during the REMD simulations. These (extended, intermediate, and compact) structures are shown along the free-energy profile over the reaction coordinate. Here, the peptide encounters various quasi-stable intermediate states between extended and compact structures. Non-synchronization between gain in potential energy and loss in conformational entropy results in a multi-basin free energy profile. We demonstrate that the alliance between the two simulation methods, ADMD and REMD, has a great synergistic effect in understanding the protein folding process.

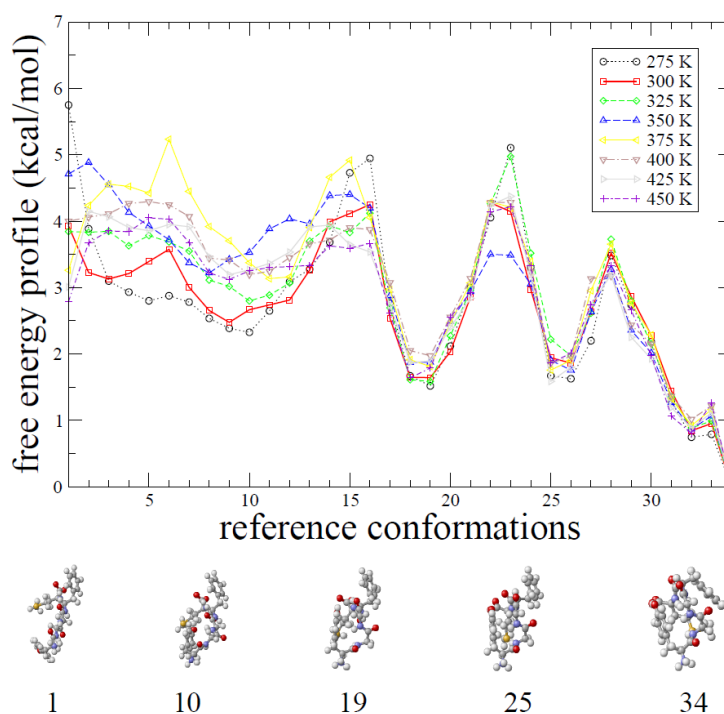
**Figure 7.** Distribution functions of the radius of gyration (a) and RMSD (b) obtained from REMD simulation are shown. Free energy functions in terms of the radius of gyration (c) and RMSD (d) are shown.



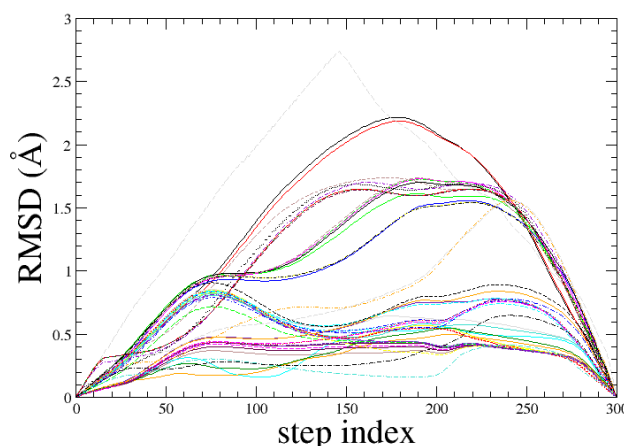
Although the goal of using ADMD approach is to find the most probable pathway out of many possibilities, we compared the lowest energy-barrier pathway to the other 49 ADMD pathways. We observe significant overlaps between the lowest energy-barrier pathway and the other 49. In order to quantify the overlap between two ADMD pathway models, we calculated a series of RMSD values between two corresponding structures from two folding models. From Figure 9, we find that calculated RMSD values are less than 1 Å over all ADMD step indices for 35 pathways out of 49. For the other 14 folding models, calculated RMSD values are below 3 Å over all step indices. This suggests that although the current ADMD approach may not guarantee to cover all pathways, the most relevant one is likely to be covered, since significant overlap is observed among the majority of 50 ADMD pathways.

One of the key questions remaining is whether the obtained free energy profile can explain populations of diverse conformations found in experiments. The fully extended structure of reference conformation 1 (or 3) is metastable with a small free energy barrier of about 1 kcal/mol as shown in Figure 8. This means that the conformational diversity especially similar to the extended structure is supported by the free energy calculation. In addition, we observe three additional metastable reference conformations, 10, 19 and 25, as intermediate conformations. These conformations represent a series of molecular compaction. Their respective structural stabilities at finite temperature are found in the free energy profile. Possibly as many as four different conformations are locally stable in addition to the final compact structure. This is consistent with NMR experimental results that suggest the existence of an ensemble of flexible structures including extended conformations in the solution structure of the peptide [25].

**Figure 8.** Free-energy profile obtained from REMD simulation based on the 34 “states” (or reference conformations) as described in text are shown. We take the reference energy of the free energy in such a way that its observed minimum is set to zero. Five reference conformations are also shown. The ground state (reference conformation 34) is favored thermodynamically and separated from the other low energy states by free-energy barriers. Reference conformation 1 is the fully extended conformation and marginally stable with a small free energy barrier of about 1 kcal/mol. Three reference conformations 10, 19 and 25 are metastable structures. For the free energy profile convergence test, we have carried out additional 50% of simulation time. The resulting two free energy profiles agree to each other within 1 kcal/mol.



**Figure 9.** The lowest energy-barrier pathway is compared to the other 49 ADMD pathways by calculating RMSD values between two corresponding conformations. We find that calculated RMSD values are less than 1 Å over all ADMD step indices for 35 pathways out of 49. For the other 14 folding models, calculated RMSD values are below 3 Å.



The relative stability of the locally-stable extended structure fluctuates as the temperature increases. In contrast, the calculated free energy profile demonstrates that, as the temperature increases, there are no dramatic changes in the profile near the compact structure. The barriers between two adjacent free energy basins at temperature 300 K is found to be about 2–3 kcal/mol, indicating that the peptide could be easily populated at two separate basins. The existence of the two well-separated stable conformations in met-enkephalin is consistent with the known affinity of met-enkephalin for distinct opioid receptors.

Experimental studies indicate that met-enkephalin does not take a single compact conformation in the aqueous solution, and an ensemble of various structures is populated [25]. An NMR study supports the presence of a turn at Gly<sup>3</sup> and Phe<sup>4</sup> in dimethyl sulfoxide. Kinoshita *et al.* found that a fully extended conformation has the highest stability in water by using the reference interaction site model theory [35]. The reference interaction site model theory is a useful tool to study solvation effects, and when it is combined with replica-exchange Monte Carlo simulations, it is shown that solvent molecules tend to extend the peptide conformation and smooth out its free-energy landscape [36]. Li and Scheraga [37] also concluded that, at room temperature, met-enkephalin in water is likely in an unfolded state. These findings, including the current study, are in qualitative agreement with NMR experimental data [25].

### 3. Theoretical Methods

#### 3.1. ADMD Method: Background and Setup Parameters

The proposed protocol is tested on a small protein, penta-peptide met-enkephalin. We used the amber95 all-atom force field with the GB/SA solvent model [38] as implemented in the TINKER package [39]. Two conformations of met-enkephalin, extended and compact, are used as the input for the ADMD simulation. The initial extended conformation was obtained by local optimization of the above energy function starting from the fully extended conformation. Similarly, the final compact conformation is obtained by local optimization of the putative ground-state structure of met-enkephalin reported in the literature [16,40–43]. Here, we note that the initial conformation is an artificial conformation, not a preferred conformation of the peptide.

ADMD simulations are typically used to study rare events such as protein folding by generating the most probable trajectories with fixed initial and final boundary conditions [9,10,44]. Here we describe the essence of the method with some details for self-containedness. In ADMD, a pre-set time interval  $[0, \tau]$  is divided into  $P$  slices of equal time length with  $\Delta = \tau/P$ . In other words, the time step is set as  $t_j = j\Delta$  with  $j = 0, 1, 2, \dots, P$ . In this study we used  $P = 300$  and  $\Delta = 8.06$  fs. The goal of ADMD is to optimize the objective function in Ref. 10 with  $3N(P - 1) = 67275$  degrees of freedom, where the total number of atoms  $N = 75$  for met-enkephalin. The atomic mass of each constituent atom (H, C, N, O, and S) is taken from its standard value and the total-energy conservation is taken into account. All atoms are treated as point particles. It should be noted that no artificial constraints on covalent bond lengths and bond angles are employed. The computational task of ADMD is to perform minimization with  $3N(P - 1)$  independent variables. Potential energy and force calculation is the most time-consuming part of the computation. To avoid various numerical difficulties, a modified action can

be used. The modified action is transformed from a functional to a function of  $P - 1$  vectors  $\{\vec{q}_j\}$ , where the index  $j$  denotes the time frame corresponding to its three-dimensional image. Here,  $j = 0$  and  $j = P$  correspond to the given initial and final images, respectively. One can define the discretized Lagrangian of the  $j$ -th temporal frame as

$$L_j = \sum_{I=1}^N \frac{m_I}{2\Delta^2} (\vec{q}_{I,j} - \vec{q}_{I,j+1})^2 - V(\{\vec{q}_j\}). \quad (2)$$

Here, the first term corresponds to the kinetic energy and  $V$  is the potential energy.  $m_I$  is the mass of the  $I$ -th atom, and  $\vec{q}_{I,j}$  is the position vector of the  $I$ -th atom at the  $j$ -th time frame. Finally, the modified action used in this ADMD becomes

$$\Phi(\{\vec{q}_j\}, E; T) = \sum_{j=0}^{P-1} L_j(\{\vec{q}_j\})\Delta + \mu_E \sum_{j=0}^{P-1} (E_j - E)^2 + \mu_K \sum_{I=1}^N \left( \langle K_I \rangle - \frac{3k_B T}{2} \right)^2, \quad (3)$$

where  $\langle K_I \rangle$  is the average kinetic energy of the  $I$ -th atom along the trajectory [9].  $E$  is the pre-set target energy of the system. A fictitious temperature  $T$  controls the kinetic energy of the system,  $\mu_E$  and  $\mu_K$  are arbitrarily large constants, and  $k_B$  is the Boltzmann constant.

The quality of a generated path can be assessed by calculating its Onsager-Machlup (OM) action [45] defined as follows;

$$S_{OM} = \frac{1}{3N(P-1)} \sum_{I=1}^N \sum_{j=1}^{P-1} \left[ 2\vec{q}_{I,j} - \vec{q}_{I,j-1} - \vec{q}_{I,j+1} - \frac{\Delta^2}{m_I} \frac{\partial V(\{\vec{q}_j\})}{\partial \vec{q}_{I,j}} \right]^2. \quad (4)$$

The term inside brackets is zero for a trajectory generated by Verlet algorithm [46]. Therefore, the OM action becomes zero for a true Verlet path in the limit of  $\Delta \rightarrow 0$ , and it can be used to estimate the proximity of a generated discrete trajectory to its ideal Newtonian trajectory, piece by piece.

With the additional  $\mu_K$  term in Equation (2), Lee *et al.* [10] demonstrated that the quality of trajectories is much improved in terms of the OM action [45] compared to that from the original action. After the final path is obtained by the action optimization process, one can estimate the quality of the path by calculating  $S_{OM}$  as the sum of  $3N \times (P - 1)$  “error variables” in Equation (3). If  $\Delta$  is small enough, a Verlet trajectory is guaranteed. The histogram of “error variables” of a typical ADMD path is shown in Figure 4. Even with relatively large  $\Delta$ , deviation from the ideal Verlet trajectory is shown to be minimal. Elber *et al.* [44] also pointed out that the distribution of “error variables” can be used as a useful indicator for assessing trajectories generated by MD simulations with non-negligible  $\Delta$ . The current ADMD approach is shown to generate high-quality trajectories with significantly small values of  $S_{OM}$  in the study of protein folding pathways [11–13,32,47] as well as systems in various condensed matter physics [48–56].

### 3.2. REMD Simulations: Setup Details

For an ideal system where the ‘exact’ force field is provided, ADMD paths represent the most probable folding trajectories at folding condition. Generally, obtaining the most probable pathway to a folded state constitutes only half of the folding story, and the other half should be concerned with thermodynamics of folding. Following the philosophy of the ADMD methodology that, when studying

a rare event such as protein folding, investigation of the most probably pathway by solving a boundary value problem is a better strategy than unguaranteed attempts for direct observation of the event from conventional MD simulations, we have designed MD simulations to study the thermodynamics of the folding along the pathway generated by ADMD. Therefore, the goal of MD simulations in this study is to estimate the free energy profile along the generated ADMD pathway. Here, we propose that this goal can be achieved by projecting all MD trajectories onto generated ADMD frames. This is based on the assumption that the folding contribution from the phase space far from a generated ADMD pathway is neither indispensable nor heavily biased toward a particular ADMD state (out of  $P + 1 = 301$  frames).

In practice, rather than dealing with all  $P + 1 = 301$  frames, we have selected a total 34 reference conformations for projection from the ADMD simulation (ADMD step indices  $j = 0, 9, 18, \dots, 270, 279, 288$  and  $300$ ). Conformations obtained from MD simulations are categorized according to their similarity with the 34 reference conformations as determined by the lowest RMSD value.

Here, we describe details about the method we employed to estimate the free-energy profile along the pathway generated by ADMD. In order to confine MD simulation trajectories in the proximity of a predetermined pathway (*i.e.*, ADMD pathway), we add the following restraint potential.

$$V_{add}(\{\vec{q}_I\}) = \begin{cases} 0 & \text{if } RMSD < 2 \text{ \AA}, \\ \frac{k}{2} \sum_{I=1}^N (\vec{q}_I - \vec{q}_{I,0})^2 & \text{otherwise.} \end{cases} \quad (5)$$

root-mean square deviation (RMSD) in Equation (4) corresponds to the lowest one out of 34 superimpositions of the current MD conformation against 34 reference conformations. We denote the closest reference conformation as  $\{\vec{q}_{I,0}\}$  where  $I$  is the atomic index of  $N = 75$  met-enkephalin atoms.

If the current MD conformation is similar to one of the 34 preset conformations within the RMSD value of 2 Å, the above restraint term does not affect the dynamics. When it deviates, *i.e.*,  $RMSD > 2 \text{ \AA}$ , the restoring harmonic restraint term in Equation (4) is applied. Here, we used  $k = 2.24 \text{ kcal/mol/\AA}^2$ . The penalty function is designed to keep the molecule within the region close to the ADMD transition pathway. We measured the portion of population affected by nonzero  $V_{add}$  from Equation (4). We find that only 1.2% of the total population was governed by nonzero  $V_{add}$ . In addition, we find that there are no specific trajectories that systematically tend to exit the transition tube. This means that neighboring reference conformations are sufficiently close to each other and most transitions observed in the REMD simulation are transitions between them. Almost all populations (98.8%) are sampled within the transition tube (defined by 2 Å range from the pathway) in the present REMD simulation. It should be noted that this penalty term does not affect the relative population along the pathway, since it is zero along the pathway within 2 Å.

The reaction coordinate chosen in this study is along the folding pathway characterized by the ADMD step index. Two successive conformations in the pathway model are quite similar to each other with a small RMSD value in the range of 0.2–0.33 Å. By means of structural clustering, we are able to determine the ensembles of the folding step including the transition state. The connectivity between successive conformations basically represents the folding process of the molecule in an overall view. Within this procedure one can calculate the population of the conformations on the reaction coordinate as the function of temperature considering multiple temperature MD simulations. This should be contrasted to *a priori* reaction coordinates typically used in conventional molecular simulations.

We have considered  $8 \times 4$  MD simulations, initially uniformly sampled from the pathway model, over eight temperatures (275, 300, 325, 350, 375, 400, 425 and 450 K). Then, based on the Metropolis criterion we attempt exchanging two conformations at neighboring temperatures [14,15]. The idea of this method is to make conformations at high temperatures available to the simulations at low temperatures and vice versa. Each MD simulation covered 60 ns while exchange attempts are tried after each MD simulation of 0.1 ns. The exchange is carried out according the conventional probability of

$$P(\text{exchange}) = \min \left[ 1, \exp \left\{ \left( \frac{1}{k_B T_i} - \frac{1}{k_B T_j} \right) (V_i - V_j) \right\} \right]. \quad (6)$$

Significant overlaps between probability density functions especially at neighboring temperatures are observed in the simulations (see Figure 6). This represents that exchanges between neighboring replicas in the current REMD simulation are properly maintained.

For an integration of the Langevin equations of motion, we used the integration time step of 1 fs. Damping constant of  $91 \text{ ps}^{-1}$  was used [57] at each temperature used. Multiple MD simulations are prepared by employing the Maxwell-Boltzmann distribution for all temperatures, ranging from  $T = 275\text{--}450$  K. We used a separate random number seed for each replica and generated random series of atomic velocities. By carrying out a single set of REMD simulations, one can obtain various thermodynamic quantities as the function of temperature [14,15]. It is shown that the REMD method can be successfully applied to the simulation of proteins with complex energy landscape. It should be noted that both ADMD and REMD simulations can be executed in a highly parallel fashion [9,10,14]. To the best of our knowledge, the current work is the first attempt to combine the ADMD method and the REMD method together.

#### 4. Conclusions

In this study, we have combined ADMD and REMD to obtain the free-energy profile along the folding pathway. Pathways are generated by solving a boundary-value problem for given initial and final conformations of a molecule in terms of ADMD formalism. By minimizing the action defined by Equation (2), Verlet-like trajectories are obtained. The quality of ADMD-generated pathways, assessed by the Onsager-Machlup action, is shown to be quite satisfactory. Furthermore, atomic motions in the pathway statistically satisfy the microscopic time reversal symmetry imposed on the Verlet trajectory. Consequently, the ADMD formulation has no limitations in describing forward as well as reverse reactions.

We choose the pathway model to be the reaction coordinate along which kinetic behavior is presumed to be described, and so the process of conformational change is characterized by the ADMD step index. In practice, we define a series of conformations, containing sequential information about the conformational change, as the reaction coordinate. This should be contrasted to the common choice of an *a priori* reaction coordinate as often used in conventional studies of molecular conformational changes.

With a generated pathway model by ADMD, the free-energy profile along the folding pathway is generated by applying restrained REMD simulations. Temperature variation of the free-energy profile over the reaction coordinate is readily evaluated by the REMD method. With the reaction coordinate of

the conformational change and its free-energy profile, the thermodynamics of conformational variation are described.

The proposed protocol is applied to a small protein, penta-peptide met-enkephalin. We show that the transition pathway of met-enkephalin between extended and compact conformations can be successfully explored by the method. Overall results are consistent with those in the literature.

It should be noted that both ADMD and REMD simulations can be executed in a highly parallel fashion. We believe that the current method can serve as a useful tool for efficient exploration of conformational space and proper characterization of protein folding dynamics. Study of conformational transitions of large and complex systems is now underway.

## Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No. 2008-0061987). We thank Korea Institute for Advanced Study for providing computing resources (KIAS Center for Advanced Computation) for this work.

## Conflict of Interest

The authors declare no conflict of interest.

## References

1. Dobson, C.M. Protein folding and misfolding. *Nature* **2003**, *426*, 884–890.
2. Fersht, A.R.; Daggett, V. Protein folding and unfolding at atomic resolution. *Cell* **2002**, *108*, 573–582.
3. Wolynes, P.G. Latest folding game results: Protein A barely frustrates computationalists. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 6837–6838.
4. Gouda, H.; Torigoe, H.; Saito, A.; Sato, M.; Arata, Y.; Shimada, I. Three-dimensional solution structure of the B domain of staphylococcal protein A: Comparisons of the solution and crystal structures. *Biochemistry* **1992**, *31*, 9665–9672.
5. Witte, K.; Skolnick, J.; Wong, C. A synthetic retrotransition (Backward Reading) sequence of the right-handed three-helix bundle domain (10-53) of protein A shows similarity in conformation as predicted by computation. *J. Am. Chem. Soc.* **1998**, *120*, 13042–13045.
6. Shaw, D.E.; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R.O.; Eastwood, M.P.; Bank, J.A.; Jumper, J.M.; Salmon, J.K.; Shan, Y.; *et al.* Atomic-level characterization of the structural dynamics of proteins. *Science* **2010**, *330*, 341–346.
7. Lindorff-Larsen, K.; Piana, S.; Dror, R.O.; Shaw, D.E. How fast-folding proteins fold. *Science* **2011**, *334*, 517–520.
8. Chodera, J.D.; Pande, V.S. Splitting probabilities as a test of reaction coordinate choice in single-molecule experiments. *Phys. Rev. Lett.* **2011**, *107*, 098102.
9. Passerone, D.; Parrinello, M. Action-derived molecular dynamics in the study of rare events. *Phys. Rev. Lett.* **2001**, *87*, 108302.

10. Lee, I.-H.; Lee, J.; Lee, S. Kinetic energy control in action-derived molecular dynamics simulations. *Phys. Rev. B* **2003**, *68*, 064303.
11. Lee, I.-H.; Kim, S.-Y.; Lee, J. Dynamic folding pathway models of  $\alpha$ -helix and  $\beta$ -hairpin structures. *Chem. Phys. Lett.* **2005**, *412*, 307–312.
12. Lee, I.-H.; Kim, S.-Y.; Lee, J. Dynamic folding pathway models of the villin headpiece subdomain (HP-36) structure. *J. Comput. Chem.* **2010**, *31*, 57–65.
13. Lee, I.-H.; Kim, S.-Y.; Lee, J. Folding models of min-protein FSD-1. *J. Phys. Chem. B* **2012**, *116*, 6916–6922.
14. Sugita, Y.; Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **1999**, *314*, 141–151.
15. Swendsen, R.H.; Wang, J.S. Replica Monte Carlo simulation of spin-glasses. *Phys. Rev. Lett.* **1986**, *57*, 2607–2609.
16. Lee, J.; Scheraga, H.A.; Rackovsky, S. New optimization method for conformational energy calculations on polypeptides: Conformational space annealing. *J. Comput. Chem.* **1997**, *18*, 1222–1232.
17. Perez, J.J.; Villar, H.O.; Loew, G.H. Characterization of low-energy conformational domains for met-enkephalin. *J. Comput. Aided Mol. Des.* **1992**, *6*, 175–190.
18. Jin, A.Y.; Leung, F.Y.; Weaver, D.F. Three variations of genetic algorithm for searching biomolecular conformation space: Comparison of GAP 1.0, 2.0, and 3.0. *J. Comput. Chem.* **1999**, *20*, 1329–1342.
19. Isogai, Y.; Nemethy, G.; Scheraga, H.A. Enkephalin: Conformational analysis by means of empirical energy calculations. *Proc. Natl. Acad. Sci. USA* **1977**, *74*, 414–418.
20. Montcalm, T.; Weili, C.; Hong, Z.; Guarnieri, F.; Wilson, S.R. Simulated annealing of met-enkephalin: low energy states and their relevance to membrane-bound, solution and solid-state conformations. *J. Mol. Struct. Theochem.* **1994**, *308*, 37–51.
21. Garbay-Jauregui, C.; Roques, B.P.; Oberlin, R.; Anteunis, M.; Lala, A.K. Preferential conformation of the endogenous opiate-like pentapeptide met-enkephalin in DMSO-d<sub>6</sub> solution determined by high field H NMR. *Biochem. Biophys. Res. Commun.* **1976**, *71*, 558–565.
22. Jones, C.R.; Gibbons, W.A.; Garsky, V. Proton magnetic resonance studies of conformation and flexibility of encephalin peptides. *Nature* **1976**, *262*, 779–782.
23. Roques, B.P.; Garbay-Jauregui, C.; Oberlin, R.; Anteunis, M.; Lala, A.K. Conformation of met-enkephalin determined by high field PMR spectroscopy. *Nature* **1976**, *262*, 778–779.
24. Khaled, M.A.; Long, M.M.; Thompson, W.D.; Bradley, R.J.; Brown, G.B.; Urry, D.W. Conformational states of enkephalins in solution. *Biochem. Biophys. Res. Commun.* **1977**, *76*, 224–231.
25. Graham, W.H.; Carter, E.S., II; Hickes, R.P. Conformational analysis of met-enkephalin in both aqueous solution and in the presence of sodium dodecyl sulfate micelles using multidimensional NMR and molecular modeling. *Biopolymers* **1992**, *32*, 1755–1764.
26. Renugopalakrishnan, V.; Rapaka, R.S.; Collette, T.W.; Carreira, L.A.; Bhatnagar, R.S. Conformational states of Leu<sup>5</sup>- and Met<sup>5</sup>-enkephalins in solution. *Biochem. Biophys. Res. Commun.* **1985**, *126*, 1029–1035.

27. Schiller, P.W. Fluorescence study on the conformation of a cyclic enkephalin analog in aqueous solution. *Biochem. Biophys. Res. Commun.* **1983**, *114*, 268–274.
28. Spirtes, M.A.; Schwartz, R.W.; Mattice, W.L.; Coy, D.H. Circular dichroism and absorption study of the structure of methionine-enkephalin in solution. *Biochem. Biophys. Res. Commun.* **1978**, *81*, 602–609.
29. D'Alagni, M.; Delfini, M.; di Nola, A.; Eisenberg, M.; Paci, M.; Roda, L.G.; Veglia, G. Conformational study of Met-enkephalin-Arg-Phe in the presence of phosphatidylserine vesicles. *Eur. J. Biochem.* **1996**, *240*, 540–549.
30. Smith, G.D.; Griffin, J.F. Conformation of Leu5 enkephalin from X-ray diffraction: Features important for recognition at opiate receptor. *Science* **1978**, *199*, 1214–1216.
31. Deschamps, J.R.; George, C.; Flippen-Anderson, J.L. Structural studies of opioid peptides: A review of recent progress in X-ray diffraction studies. *Biopolymers* **1996**, *40*, 121–139.
32. Lee, I.-H.; Kim, S.-Y. Dynamics of the neuropeptide met-enkephalin by using action-derived molecular dynamics. *J. Korean Phys. Soc.* **2008**, *53*, 1764–1769.
33. Kearsley, S.K. An algorithm for the simultaneous superposition of a structural series. *J. Comput. Chem.* **1990**, *11*, 1187–1192.
34. Coutsiias, E.A.; Seok, C.; Dill, K.A. Using quaternions to calculate RMSD. *J. Comput. Chem.* **2004**, *25*, 1849–1857.
35. Kinoshita, M.; Okamoto, Y.; Hirata, F. Solvation structure and stability of peptides in aqueous solutions analyzed by the reference interaction site model theory. *J. Chem. Phys.* **1997**, *107*, 1586–1599.
36. Mitsutake, A.; Kinoshita, M.; Okamoto, Y.; Hirata, F. Combination of the replica-exchange Monte Carlo method and the reference interaction site model theory for simulating a peptide molecule in aqueous solution. *J. Phys. Chem. B* **2004**, *108*, 19002–19012.
37. Li, Z.; Scheraga, H.A. Structure and free energy of complex thermodynamic systems. *J. Mol. Struct. (Theochem.)* **1988**, *179*, 333–352.
38. Qiu, D.; Shenkin, P.S.; Hollinger, F.P.; Still, W.C. The GB/SA continuum model for solvation. A fast analytical method for the calculation of approximate Born radii. *J. Phys. Chem. A* **1997**, *101*, 3005–3014.
39. Ponder, J.W.; Richard, F.M. An efficient Newton-like method for molecular mechanics energy minimization of large molecules. *J. Comput. Chem.* **1987**, *8*, 1016–1024.
40. Eisenmenger, F.; Hansmann, U.H.E. Variation of the energy landscape of a small peptide under a change from the ECEPP/2 force field to ECEPP/3. *J. Phys. Chem. B* **1997**, *101*, 3304–3310.
41. Hayryan, S.; Hu, C.-K.; Hu, S.-Y.; Shang, R.-J. Multicanonical parallel simulations of proteins with continuous potentials. *J. Comput. Chem.* **2001**, *22*, 1287–1296.
42. Vengadesan, K.; Gautham, N. Conformational studies on enkephalins using the MOLS technique. *Biopolymers* **2004**, *74*, 476–494.
43. Zhan, L.; Chen, J.Z.Y.; Liu, W.-K. Conformational study of met-enkephalin based on the ECEPP force fields. *Biophys. J.* **2006**, *91*, 2399–2404.
44. Elber, R.; Cárdenas, A.; Ghosh, A.; Stern, H.A. Bridging the gap between reaction pathways, long time dynamics and calculation of rates. *Adv. Chem. Phys.* **2003**, *126*, 93–129.
45. Onsager, L.; Machlup, S. Fluctuations and irreversible processes. *Phys. Rev.* **1953**, *91*, 1505–1512.

46. Verlet, L. Computer “Experiments” on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. *Phys. Rev.* **1967**, *159*, 98–103.
47. Lee, I.-H.; Kim, S.-Y.; Lee, J. Dynamics of conformational isomerization in alanine dipeptide and valine dipeptide. *J. Korean Phys. Soc.* **2005**, *46*, 601–605.
48. Kim, Y.-H.; Lee, I.-H.; Chang, K.J.; Lee, S. Dynamics of fullerene coalescence. *Phys. Rev. Lett.* **2003**, *90*, 065501.
49. Lee, I.-H.; Kim, H.; Lee, J. Dynamic pathway model for the formation of C<sub>60</sub>. *J. Chem. Phys.* **2004**, *120*, 4672–4676.
50. Lee, I.-H.; Jun, S.; Kim, H.; Kim, S.Y.; Lee, Y. Adatom-assisted structural transformations of fullerenes. *Appl. Phys. Lett.* **2006**, *88*, 011913.
51. Pendurti, S.; Jun, S.; Lee, I.-H.; Prasad, V. Cooperative atomic motions and core rearrangement in dislocation cross slip. *Appl. Phys. Lett.* **2006**, *88*, 201908.
52. Kim, S.Y.; Lee, I.-H.; Jun, S.; Lee, Y.; Im, S. Coalescence and T-junction formation of carbon nanotubes: Action-derived molecular dynamics simulations. *Phys. Rev. B* **2006**, *74*, 195409.
53. Kim, S.Y.; Lee, I.-H.; Jun, S. Transition-pathway models of atomic diffusion on fcc metal surfaces. I. Flat surfaces. *Phys. Rev. B* **2007**, *76*, 245407.
54. Kim, S.Y.; Lee, I.-H.; Jun, S. Transition-pathway models of atomic diffusion on fcc metal surfaces. II. Stepped surfaces. *Phys. Rev. B* **2007**, *76*, 245408.
55. Lee, Y.; Han, J.; Lee, I.-H.; Im, S. Mobility of a 5|7 defect in carbon nanotubes. *Nanotechnology* **2011**, *22*, 105707.
56. Lee, I.-H.; Kim, S.Y.; Jun, S. An introductory overview of action-derived molecular dynamics for multiple time-scale simulations. *Comput. Methods Appl. Mech. Eng.* **2004**, *193*, 1633–1644.
57. Zagrovic, B.; Sorin, E.J.; Pande, V.  $\beta$ -hairpin folding simulations in atomistic detail using an implicit solvent model. *J. Mol. Biol.* **2001**, *313*, 151–169.