Energy Landscape of Met-Enkephalin and Leu-Enkephalin Drawn Using Mutually Orthogonal Latin Squares Sampling

K. Vengadesan and N. Gautham*

Department of Crystallography and Biophysics, University of Madras, Chennai 600 025, India

Received: January 30, 2004; In Final Form: April 30, 2004

Mapping the energy surface is an important step in understanding the protein folding process. However, it requires identification of all stationary points of the system, or at least a large number of them. Locating the full set of stationary points even for Met-enkephalin is intractable by normal sampling procedures that have been proposed to search the energy surface. We have recently suggested an enhanced combinatorial sampling technique that uses mutually orthogonal Latin squares to quickly identify the complete set of low potential energy structures of a molecule. In this paper, we have used this technique to identify 1500 low potential energy structures each for Met-enkephalin and Leu-enkephalin. We have then used a clustering procedure to reduce the sample to a set of unique conformations. Some of these have been identified earlier by other computational simulations as global energy minimum structures. Others structures in the unique set have been observed in experiments, specifically X-ray crystallography. We have used the unique sets to visualize the entire potential energy landscape as two-dimensional projections and as minimum energy envelopes. The potential energy surfaces of these peptides resemble a funnel that is broader, as well as possessing a larger number of minima that are more widely separated, than the picture demonstrated generally in the past. We are able to locate the theoretical and experimental structures on this landscape. In addition, we are able to identify a set of structures that may represent intermediate steps on the folding pathway of these two peptides.

Introduction

The folding of peptides and proteins to form their native structure is now commonly treated as a conformational search on a multidimensional funnel-shaped energy landscape.[1−6] The funnel is defined as a deep depression on the multidimensional energy surface, which is centered over the polypeptide system's global energy minimum (i.e., the folded native state). The funnel is wide at the top and then narrows until it reaches its lowest point. Once a system enters the mouth of the funnel it is led through a multitude of pathways down to the native structure, which is assumed to be at the bottom of the funnel.[7] The idea of such folding landscapes comes from the understanding that a large energy difference between the unfolded state and the native state provides sufficient bias in the conformational space, such that both the energy and the entropy decrease as protein folds to its lowest energy (native) state.[8]

Recently many research groups have used all-atom calculations to examine the funnel hypothesis for peptides. They have used a variety of different procedures to sample the energy surface as well as to analyze the sample. A few examples follow. Sheinerman and Brooks[9] used molecular dynamics simulations in conjunction with importance sampling to map the energy surface of a small α/β protein. The free energy surface was visualized by plotting it as a function of a reaction coordinate defined by the number of native contacts. Hansmann, Okamoto, and Onuchic[3] mapped the energy landscape of Met-enkephalin using a generalized-ensemble version of the Monte Carlo simulation technique. They visualized their results by graphing the free energy as a function of three order parameters, namely the volume, overlap with the ground-state structure, and overlap with the next higher energy structure. Klepeis and Floudas[10] have used combination of the deterministic global optimization approach for low energy ensembles and eigenmode searches for transition states and minimum−transition−minimum triples to characterize the energy surface of Met-enkephalin by rate disconnectivity graphs. Alves and Hansmann[11] again used generalized-ensemble Monte Carlo simulations to calculate the free energy landscapes of two peptides (Met-enkephalin and poly-alanine) and map them as a function of temperature, and as a function of overlap with the ground-state structure. Levy and Becker[5] studied the energy landscape of three conformationally constrained hexapeptides. For each molecule, they sampled 500 conformations from a high-temperature molecular dynamics trajectory. These sampled structures were annealed down to 300 K and minimized. They used principal coordinate analyses to reduce the dimensionality of the conformational space. The potential energy surface (PES) of each molecule was visualized as a minimum energy envelope, plotted as a function of the first few principal coordinates that captured between 32% and 79% of the variance in the data for the different peptides studied. Ghosh, Elber, and Scheraga[12] carried out a study of 130 folding pathways of the 60-residue, three-helix bundle fragment B of staphylococcal protein A using the stochastic difference equation. They used the number of native hydrogen bonds and contacts and the radius of gyration as parameters to plot "free energy profiles". Sanbonmatsu and García[13] used replica exchange molecular dynamics in explicit aqueous solution to study Met-enkephalin, and visualized the energy landscape by plotting the first principal axis against temperature. Arkin and Çelik[14] used multicanonical molecular dynamics to obtain a three-dimensional topographic picture of the energy

* Address correspondence to this author. E-mail: gautham@unom.ac.in, crystal@giasmd01.vsnl.net.in.

10.1021/jp049581v CCC: $27.50
Published on Web 06/25/2004

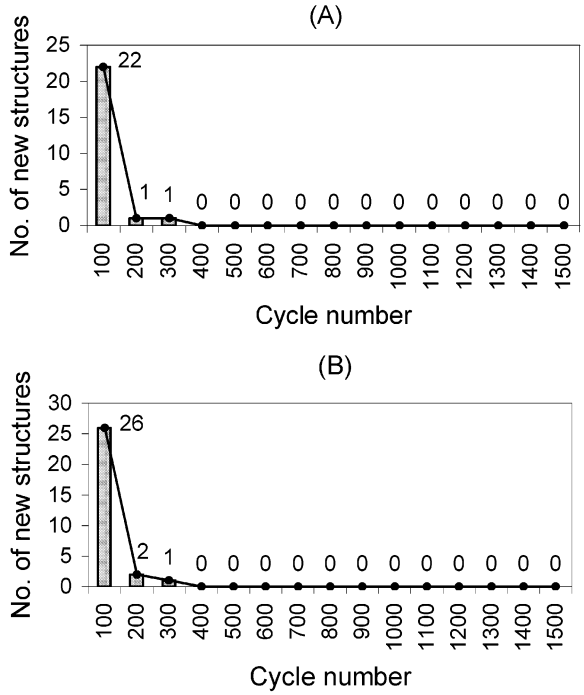

Figure 1. Number of mutually dissimilar low potential energy structures identified versus number of structures generated using MOLS. After the first 300 structures, no new structures are discovered, showing that all possible low energy conformations of the molecule have already been identified. (A) Met-enkephalin; (B) Leu-enkephalin.

landscape of Met-enkephalin. They plotted histograms with respect to energy (temperature) and the order parameter, taken as overlap with the reference global energy minimum (GEM) structure. Evans and Wales[15] applied the discrete path sampling technique to map the free energy landscape of Met-enkephalin. Analysis using disconnectivity graphs showed two funnels in the landscape.

The results of all these computations showed that, in general, and in keeping with theoretical expectations, the overall shape of the energy surface resembles a funnel. However, the surface has a large number of minima and transition points, and there are numerous states with comparable energies at significant conformational distances from each other. Folding on this landscape proceeds via a large number of pathways, some of which may lead to trapping in local minima.

We have chosen this set of investigations as examples because they illustrate the variety in the all-atom methods used to sample the energy landscape. To a greater or lesser degree, the maps generated by these methods validate the funnel hypothesis and, additionally, yield information about other features of the landscape, such as multiple minima. However, all the techniques we have considered above may be classified very broadly as variations of either molecular dynamics (MD) or Monte Carlo (MC) simulations, and there may still be some question whether the multidimensional energy space has been completely sampled in an unbiased manner. In MD simulations, bias is introduced by the choice of the initial conformation of the molecule, while MC simulations sample random points on the energy surface. Even though additional variety has been introduced in the simulations by the different potential functions that have been used, none of the GEMs identified by the calculations correspond to the experimentally observed crystal structures for the enkephalins.[16−19]

We have developed a new sampling technique, which uses mutually orthogonal Latin squares (MOLS) to sample the

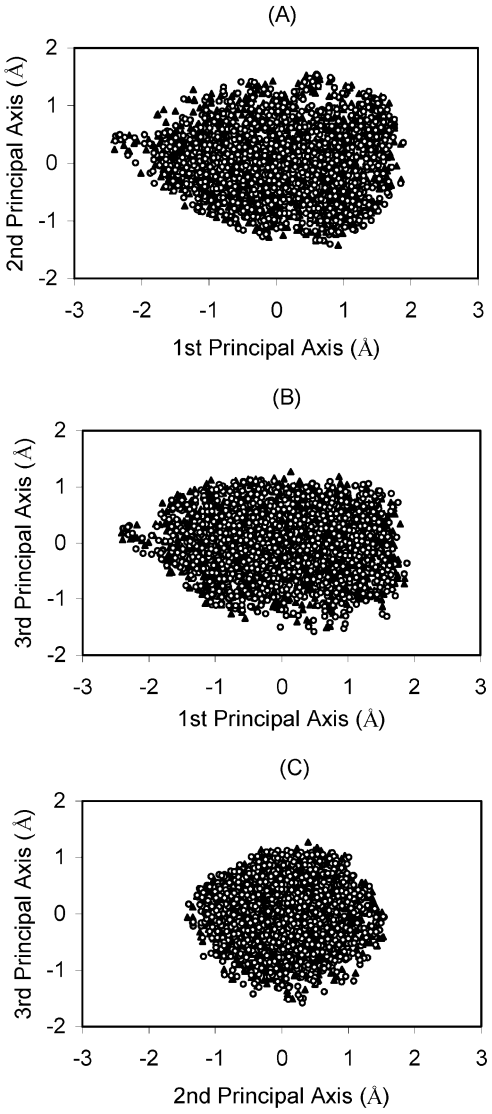

Figure 2. Two different conformation samples (consisting of 1500 conformations each) of Met-enkephalin simultaneously projected onto three two-dimensional planes (A−C), each figure representing a pair from the first three principal axes. Open circles represent conformations from the first conformation sample; filled triangles represent conformations from the second. The complete overlap between the two conformations samples is apparent and indicates that the conformation sample covers the entire available conformation space.

PES and identify low potential energy structures in a demonstrably exhaustive fashion. The procedure may be thought of as a search for the lowest potential energy structures of the molecule from a "frozen ensemble" of all conformations. We have used this method to identify the complete set of low potential energy structures of small peptides such as the enkephalins. We have shown that this set includes not only the GEM of other simulations and the structures obtained by experiments, but also some well-folded structures that were not earlier identified by any other method, experimental or computational.[20,21] In this paper, we report the use of this technique to visualize the potential energy landscapes of two neuropeptides, Met-enkephalin and Leu-enkephalin. The term energy surface (or landscape) is sometimes used for the multidimensional potential energy surface that underlies the molecule's conformation space, and sometimes for its free energy profile. In the present study our focus is on the potential energy surface.

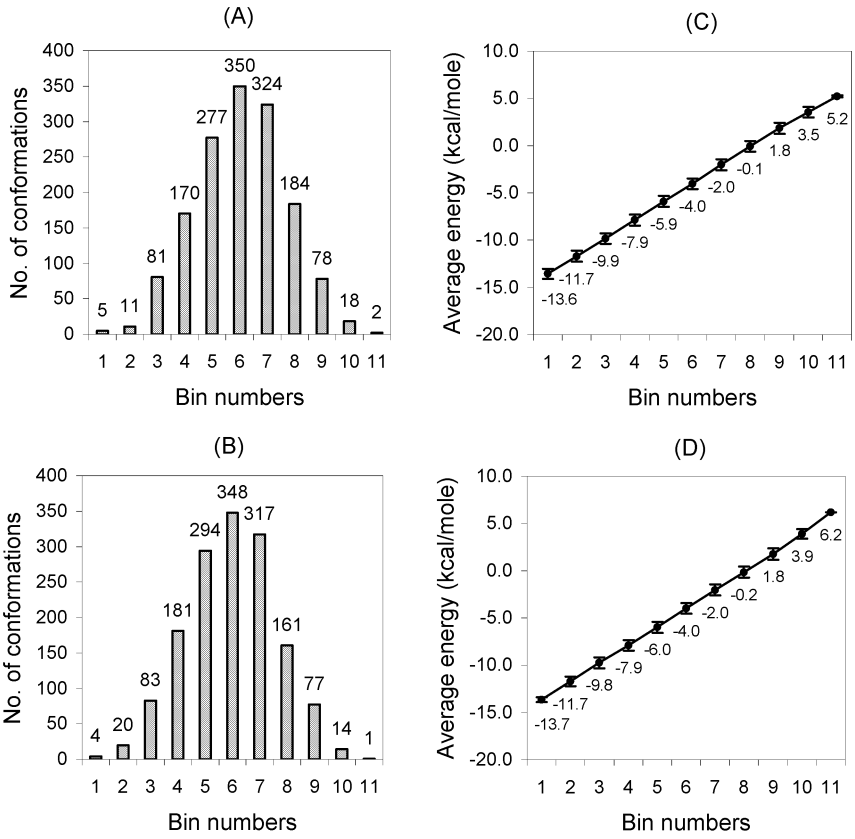

Figure 3. Distribution of the two Met-enkephalin conformation samples. Number of conformations per energy bin of size 2.0 kcal/mol for conformation sample 1 (A) and sample 2 (B). Average energy with standard deviation of each energy bin for sample 1 (C) and sample 2 (D). The complete overlap of the two sample energy profiles is apparent.

Methods

The MOLS Technique. The MOLS technique as used to sample peptide conformational space has been described in detail previously,[20,21] and we will not repeat the detailed description here, but only reiterate some of ideas underlying the algorithm. Mutually orthogonal Latin squares are used in experimental design in diverse fields such as agriculture and clinical testing.[22,23] For example, in agriculture MOLS are used to design experiments, the results of which are then analyzed to identify the particular combination of factors that leads to the maximum yield for a given crop. In general, in such cases, it is not possible to perform an experiment to probe the effect of every possible combination of the factors, since the number of such experiments would be astronomical. MOLS are therefore used to sample the experimental space in a regular and rational manner, such that the dimensionality of the experimental space is drastically reduced, without largely compromising the information content. A single Latin square of order M allows one to investigate the effect of setting three factors each at M different values in only M^2 trials, instead of M^3 trials that would be required for a complete and exhaustive sampling. A set of N mutually orthogonal Latin squares of order M allows one to investigate the effect of setting N factors each to M different values, again in only M^2 (instead of M^N) trials. The sampling is complete and exact if the factors are independent of each other.[23] In applying this technique to analyze peptide conformation, we assume that the search for a "good", i.e., low potential energy conformation may be carried out by sampling the multidimensional conformational space systematically. Thus, we treat this as a problem in experimental design. The N rotatable torsion angles in the peptide are the factors, each being allowed to assume M values in the range 0−360°. The total search space thus consists of M^N different sets of values, each set specifying one possible conformation for the peptide. The use of MOLS allows us to map the entire search space in only M^2 "experiments". These experiments consist of setting the torsion angles to the M^2 different sets of values as specified by the MOLS design, and evaluating the potential energy for each conformation. To do this, we used the ECEPP/3 force field[24] (setting the dielectric constant equal to 2). We did not include any explicit solvation model. However, as noted by other authors,[3] the force field reflects implicit solvent effects, since the parameters were calculated from crystal structures of amino acids, and have been refined to adjust rotational barriers close to experimental values. After one cycle of M^2 evaluations, the map so obtained is "inspected" by a procedure described elsewhere.[20,21] This allows us to identify one low potential energy structure, which is then minimized by gradient minimization to bring it to the nearest local minimum on the PES. If the factors, i.e., the torsion angles, were independent of each other, in that they each independently affect the potential energy, the low energy structure obtained would be the GEM. This, however, is clearly not the case, and we have shown[20,21] that the low energy structure in one cycle is only one of the many minima on the energy surface. To locate another minimum, we perform another cycle of calculations, again selecting M^2 points in the conformational space using a different set of N MOLS. For N factors with M values each, there are $(M!)^N$ different ways of choosing a set of N MOLS. Using any one of them as the basis for one cycle of calculations would lead to a minimum. Of course, this does not mean that there are $(M!)^N$ minima for the molecule. Obviously most of the cycles would lead to the conformations already discovered in previous cycles. We generated 1500 structures each for Met-enkephalin and Leu-enkephalin. (The MOLS procedure is fast.

The calculations to identify the entire set of 1500 structures took only about 15 h for Met-enkephalin and 14 h for Leu-enkephalin on a single Pentium IV 1.8 GHz processor.)

In Figure 1, we have plotted the number of new structures discovered as a function of the cycle number for the enkephalins. For each cycle of calculations leading to one minimum, we have chosen the set of N MOLS randomly as one of the $(M!)^N$ possible ways. Met- and Leu-enkephalin were modeled as described below, and the conformations were considered the same if they belonged to same cluster, as determined by the SCAR clustering program.[25] We obtained 24 and 29 significant clusters for Met-enkephalin and Leu-enkephalin, respectively. The figure shows that 300 cycles are sufficient to identify all the significant low energy clusters for both Met- and Leu-enkephalin. For Met-enkephalin there were 22 clusters that had at least one structure in the first 100 cycles. After 1500 cycles, the number of members in these clusters ranged from 8 to 100, with an average of 59. The single cluster that appeared in the second 100 cycles had a final size of 61 members, and a cluster with final size of 5 members appeared in the third 100 cycles. For Leu-enkephalin, there were 26 clusters, with a range of final size from 30 to 104, with average 58, in the first 100 cycles. There were two clusters with final cluster sizes of 35 and 70 in the second 100 cycles and a cluster with final size of 4 in the third 100 cycles. Clearly, therefore, sampling the conformational space using MOLS leads to an exhaustive identification of the energy minima. This statement is supported in three further ways. First, as described in detail elsewhere,[20,21] using the MOLS method on Met-enkephalin and Leu-enkephalin we have been able to identify not only all previously identified low energy structures, as well as all the experimentally determined structures, but also well-folded low energy conformations that have not been reported earlier. This is despite the fact that the potential energy function we have used does not represent all the different experimental conditions, nor is it the same as the potential functions used in the other theoretical studies. Considering the large amount of structural studies that enkephalins have been subjected to, this indicates that the MOLS method performs a more comprehensive search. It also indicates that the ECEPP/3 function we have used is sufficient to approximately model a variety of different experimental conditions. Second, the sample overlap procedure as proposed by Levy and Becker[5] was used to check whether the sampling procedure is exhaustive. According to this, two conformation samples of the same system are generated by two different sampling protocols (e.g., different initial conditions or different methods). If the two samples overlap and occupy the same conformation space, then the sampling procedure is exhaustive. The evaluation of the overlap between two samples can be carried out with the aid of principal component projections through a joint projection onto the same low-dimensional principal subspace. The procedure is discussed more fully below, in the context of visualization of the energy landscape. Here, we see from Figure 2 that two samples of 1500 low energy structures each of Met-enkephalin, generated as described above, occupy almost exactly the same regions of the principal subspace. The result is the same for Leu-enkephalin. This indicates that the 1500 structures for each molecule completely represent all the minima on the PES. Third, as seen in Figure 3 the energy profiles of the two groups of sampled conformations are also very similar. We are confident, therefore, that with the first 1500 samples, we have completely sampled the conformational space, and have accounted for all the minima (i.e., lowest potential energy conformations). We use these minima to visualize the energy landscape.

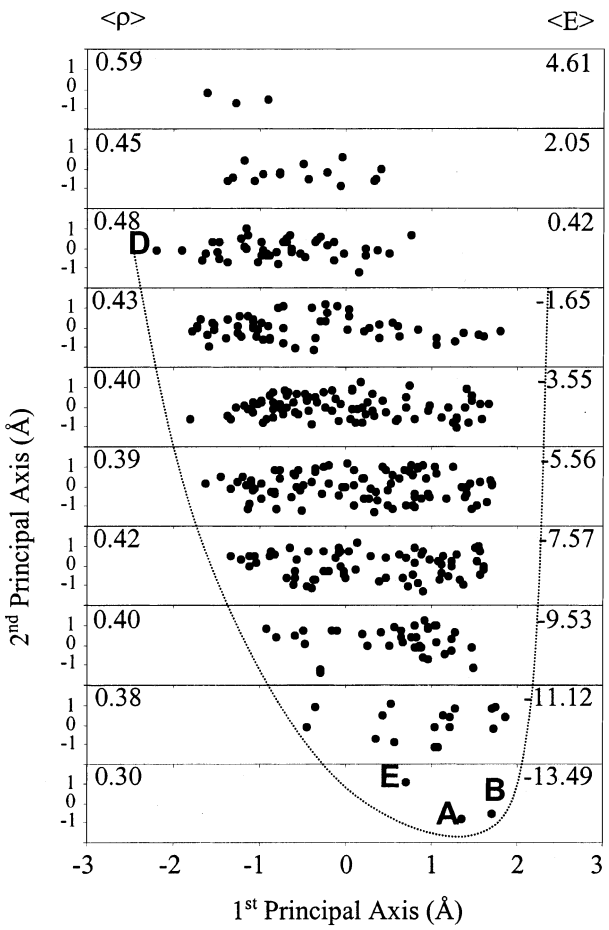

Figure 4. Principal two-dimensional projection of the conformation space of Met-enkephalin. Each rectangular slab is the plane corresponding to the two principal coordinate axes of the molecule. On each, only conformations with energy in the given energy ranges are plotted. Thus, the lowest slab has conformations in the lowest energy bin, and so on. The average energy of each slab is indicated at the top-right corner and the average value of the order parameter ρ is indicated at the top-left corner of each rectangle. The dashed line helps to visualize the underlying basin topography.

Visualizing the Energy Landscape. We have used the method suggested by Levy and Becker[5] to visualize the energy landscapes of the two molecules. This is a complex task, since even relatively small molecules have very large-dimensional conformation spaces. However, in practice, a much smaller number of dimensions are sufficient to capture the essential information required for energy landscape cartography. This is achieved by projecting the full multidimensional space on an appropriately low dimensional subspace. Reducing the dimensionality of multidimensional conformational spaces can be achieved by principal component analysis (PCA). The PCA applied in this study is the so-called principal coordinate analysis (PCoorA[5]). Briefly, PCoorA operates on the $n \times n$ matrix known as the "distance matrix", reflecting the relationships between conformations. The distance between any two conformations is measured as the root-mean-square difference in the atomic positions. The distance matrix is transformed into a centered matrix, which is then diagonalized. The resulting eigenvalues are normalized to give the percentage of the projection of the original distribution on the new set of coordinates. For Met-enkephalin the first three normalized eigenvalues were 23.8%, 9.3%, and 8.4% of the total value, indicating that the accuracy

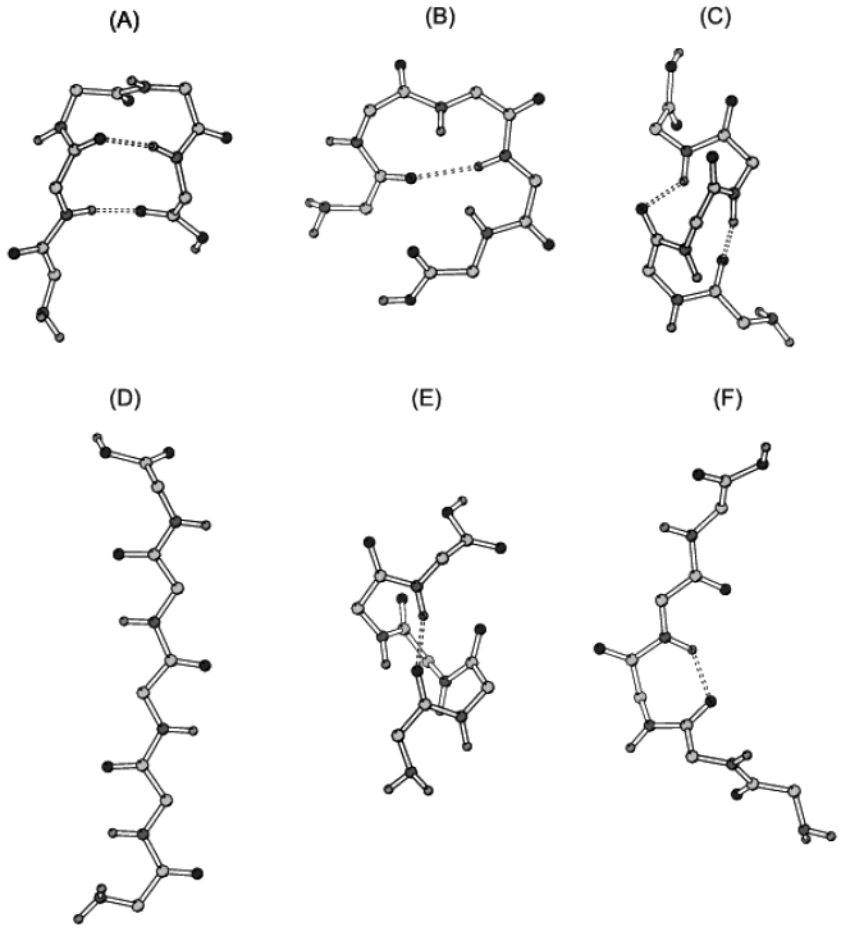

Figure 5. Classification of experimental and theoretical conformations of enkephalins. The structure shown is the one with the lowest energy in its group. (A) Single folded conformation (with 2 → 5 hydrogen bonds) centered at Gly^3-Phe^4. (B) Single folded conformation (with a 1 → 4 hydrogen bond) centered at Gly^2-Gly^3. (C) Double folded conformations (with 1 → 4 and 2 → 5 hydrogen bonds). (D) Extended conformations. (E) Helical conformations (with a 1 → 5 hydrogen bond), also called α-turns. (F) γ-Turn conformations (with a 2 → 4 hydrogen bond) centered at Gly^3.

of joint three-dimensional (3D) projection of the data points in this subspace is about 42.5%. For Leu-enkephalin, the accuracy is 42.9%. The eigenvectors, scaled by their corresponding eigenvalues, give the coordinates of the original data points in the new axes. The best two principal coordinates are used to view the energy landscape. The landscape is plotted as the minimum energy envelope in the two-dimensional (2D) principal subspace, since the MOLS method, as implemented currently, allows us to identify all the minima, but not the transition points.

As discussed above, the 1500 structures in the sample do not all have different conformations, and it would be redundant to use all of them to construct the energy landscape. We have therefore clustered together similar conformations and selected a representative conformation from each cluster. Here we have used a more stringent criterion of similarity in the structure and have taken all the backbone atoms into consideration, not just the C^α atoms as in the SCAR algorithm. We have used the algorithm suggested by Kříž et al.,[26] with a distance cutoff of 1.0 Å. After the clustering, we obtained 427 different structures for Met-enkephalin. To verify that the clustering had not affected the quality of the samples, PCoorA was used to compare the conformation coverage before and after clustering. For the full sample the joint 2D projection using the first two principal coordinates is 33.1% before clustering and 34.3% after clustering. For the joint 3D projections using the first three principal coordinates, the coverage is 42.5% and 42.5%, before and after clustering, respectively. This indicates that the reduction does not adversely affect the quality of the sample. For Leu-enkephalin, the clustering procedure reduced the 1500 conformations to 394. The percentages of the joint 2D and 3D projections are 34.0% and 42.9%, respectively, before the reduction and 34.0% and 42.7% after the reduction.

Molecular System. The model we have used for the two enkephalins has been described earlier.[21] We repeat it here for completeness. Met-enkephalin and Leu-enkephalin are pentapeptides with sequences H-Tyr-Gly-Gly-Phe-Met-OH and H-Tyr-Gly-Gly-Phe-Leu-OH, respectively. We have modeled the two peptides as each consisting of 13 degrees of freedom. Of these, 10 were the backbone Ramachandran angles ϕ and ψ. The range for each angle was taken to be 0−360° in steps of 10°. The side chain conformations for the three non-glycyl residues were chosen from the library of rotamers devised by Tuffery et al.,[27] and therefore assigned one degree of freedom each, giving a total of 13 variable parameters to describe each molecule.

Results

The two peptides have been extensively studied both experimentally and theoretically. Even though many theoretical studies showed a single GEM for Met and Leu-enkephalin,[28] experimental studies (specifically X-ray crystallography) observed many conformations,[16−19] none of them the same as the theoretical GEM. In general, conformations of enkephalins observed by experimental studies can be classified into four

groups based on the hydrogen bonding pattern: single folded conformations with a 2 → 5 hydrogen bond folded at Gly^3-Phe^4 (A), single folded conformations with a 1 → 4 hydrogen bond folded at Gly^2-Gly^3 (B); double folded conformations (C) with both the above folds, and extended conformations (D). Since there is no hydrogen bond present in group D, we used the Lewis criterion.[29] According to this, an extended conformation has a distance greater than 10 Å between C^α atoms of the residues i and $i + 3$. For both enkephalins, the theoretical GEM structure has also been identified in solution by spectroscopy[30,31] and corresponds to a group A structure. The crystal structure of Met-enkephalin[18] belongs to group D. One more folded structure has been identified for this peptide by theoretical studies.[13,32] It cannot be classified into any of the above experimental determined groups of structures, and is called E. For Leu-enkephalin, three different structures have been identified by X-ray crystallography. They are a single folded structure belonging to group B,[16] a double folded conformation belonging to group C,[19] and an extended conformation belonging to group D.[17,18] All these structures have been identified in the present calculations and are represented on the energy landscape. In each case, the structure with the lowest energy has been chosen to represent the group.

Minimum-Based View of the Energy Landscape. We have chosen two different ways of representing the energy landscape.[5] Figure 4 shows a quantitative map of the landscape for Met-enkephalin. In this figure, the structures are divided according to energy into 10 bins of width 2.0 kcal/mol each. The structures in each bin are projected onto a rectangular two-dimensional plane with the first and second principal coordinates making up the two axes. For ease of visualization, the two axes are not on the same scale. However, each rectangular slab is a full 2D principal coordinate projection, on which all those conformations that have energy within a given range are indicated. The projections are then stacked together. The funnel-like structure of the landscape, especially with respect to the first principal axis, is clearly seen. The second principal axis has almost the same narrow range between −2.0 and 2.0 Å in all the energy slabs. The third and higher principal coordinates (not shown) also have narrow ranges at all energies, indicating that most important folding coordinate is the first principal axis.

The lowermost energy slab in Figure 4 includes the three lowest energy conformations. They belong to groups A, B, and E described above. The lowest energy structure A is a type II′ β-turn centered at Gly^3-Phe^4, as shown in Figure 5A. The second lowest energy conformation B has 1 → 4 fold stabilized by a hydrogen bond between Tyr^1(CO) → Phe^4(NH) as shown in Figure 5B. The third lowest energy conformation E is an α-turn with a hydrogen bond between Tyr^1(CO) → Met^5(NH), as shown in Figure 5E. All three structures are consistent with other theoretical studies. In particular, the 2 → 5 fold in structure A has been widely observed as the GEM conformation.[3,10,14,15,28,32,34] The distances between the three structures A, B, and E along the first principal coordinate are less than 1.0 Å. However, the main chain rmsd (root-mean-square deviation in atomic positions after least-squares superposition of the two molecules) between them is greater than 1.8 Å (A/B = 1.8 Å; A/E = 2.4 Å; B/E = 2.5 Å). The crystal structure conformation D (Figure 5D) is marked in the eighth energy bin. No extended structures were found with energies below −0.38 kcal/mol. This is may be because the extended form in the solid state is stabilized by a large number of intermolecular hydrogen bonds, whereas all calculations, including the present ones, have been carried out on a single, isolated molecule.

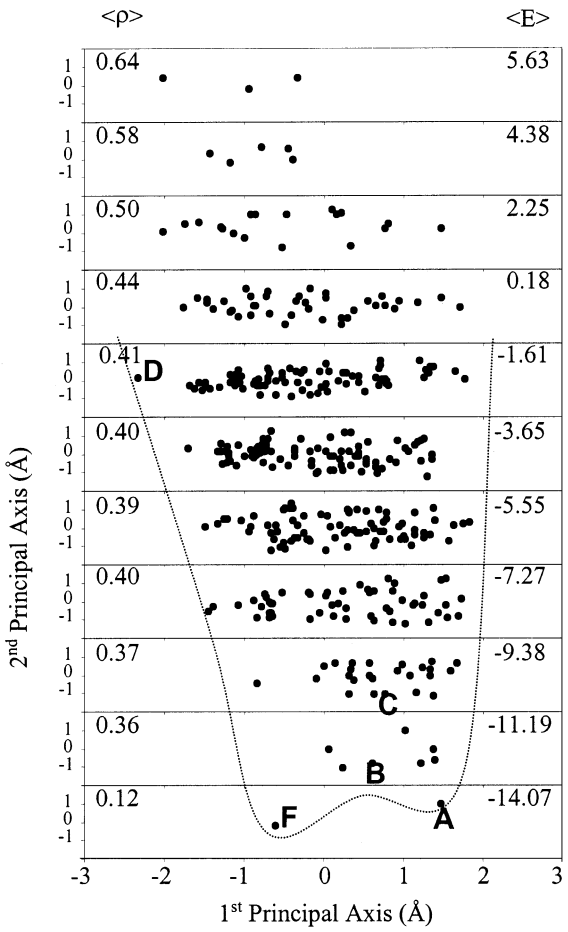

Figure 6. Principal two-dimensional projection of the conformation space of the Leu-enkephalin. (See caption of Figure 4 for details.)

Figure 6 shows the quantitative map for Leu-enkephalin. Again, we have marked the known experimental and theoretical minima. The figure shows that a change of one amino acid in the structure from methionine to leucine has a marked effect on the energy landscape. Now there are at least two equally deep "wells", separated by a large conformational distance. One corresponds to structure (A) with a type II′ β-turn centered at Gly^3-Phe^4. This is the theoretical GEM structure as determined by previous studies.[28,33] The structure in the second well has a 2 → 4 inverse γ-turn at Gly^3 as shown in Figure 5F. This structure has not been observed earlier experimentally or theoretically. However, it is a predominant and stable conformation observed in previous computational studies on the other peptide, Met-enkephalin.[34,35] The single folded conformation B, which is a 1→ 4 β-turn, is in the second energy bin. The conformation class C with a 3_{10} helix occurs in the third lowest energy bin. Both these have been observed in crystals of Leu-enkephalin.[16,19] B and C are very close to each other along the first principal axis. No extended structures were found with energy less than −0.92 kcal/mol. The extended conformation (D), seen in other Leu-enkephalin crystals,[17,18] is far from all other structures. This is the same as for Met-enkephalin.

An alternative representation of the landscape, also based on local minima only, is the minimal energy envelope. The energy envelope underlying the 2D principal projection is calculated and presented as a 3D topographical map. For Met-enkephalin, this is shown in Figure 7. It shows a broad basin, with a "well" at one end that corresponds to structure A. The bottom of the basin slopes toward this well and exhibits a significant amount of roughness. Structures A and B are closely placed, but

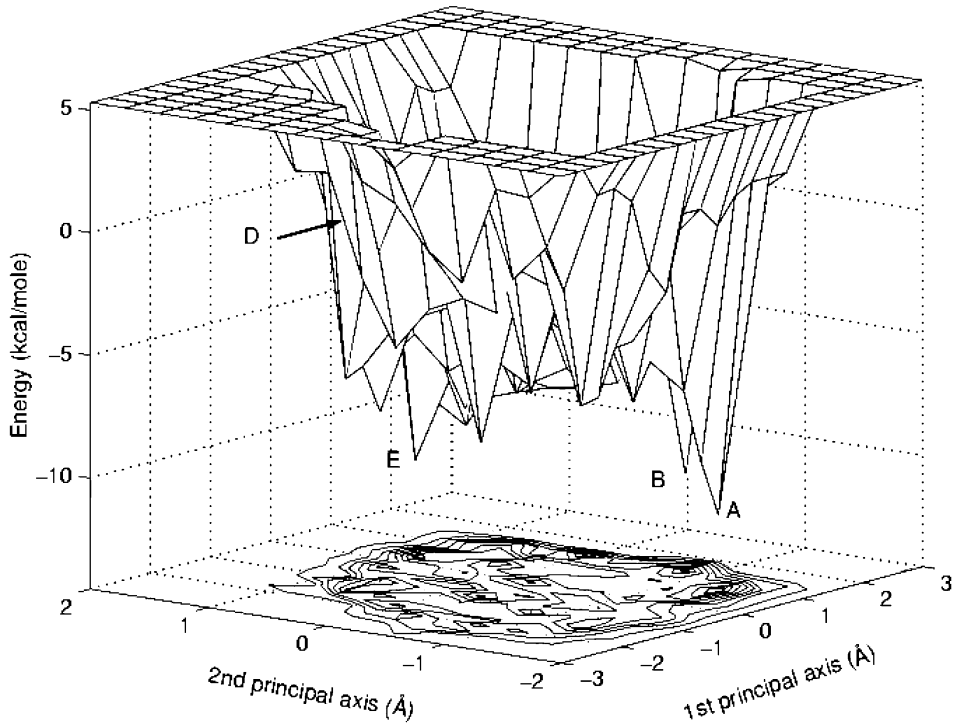

Figure 7. Energy landscape obtained by the minimal energy envelope procedure for Met-enkephalin. The two principal axes indicate conformational similarity, and the vertical axis reflects the relative energy.

structure E is far from them. This may represent a "nonnative" minimum energy trap. The experimental structure D is far from all others, some distance up the slope. We must caution, however, that roughness of the basin is a probable consequence of the method we use to obtain the map. In the first place, we have sampled only the minima and have no information regarding the transition points between one structure and another. Wales and co-workers[4] have used a "eigenvector-following" method to identify transition points, but such a methodology is not possible here. Second, this is only a partial projection and, to clarify the picture, the energy surface was smoothed; i.e., the peaks (but not the troughs) are an artifact of the (ΔX, ΔY) parameters (interval along the axes) used in drawing the map.[5] As the grid mesh size is made finer, the resulting surface becomes more detailed, but also rougher. We have chosen the value of this parameter to best illustrate the features discussed here. This means, however, that any potential energy barriers (or the lack of them) between the different minima are not well depicted here.

In Figure 8, we have given the minimal energy envelope representation of the energy landscape for Leu-enkephalin. Again, we have marked the known experimental and theoretical minima. The two equally deep wells are clearly seen. As compared to the energy surface of Met-enkephalin there are no other major changes, especially in the relative positions of A and D.

Other Order Parameters. Apart from the principal coordinates, an order parameter Q, which reflects the fraction of native contacts or conformations in any given conformation, is also frequently used as a one-dimensional reaction coordinate to visualize the multidimensional energy landscape.[2,5,6] We used the Hamming distance[36] as the order parameter. For conformation i the order parameter ρ is calculated as

$$\rho(i) = \frac{1}{nx}\sum_k \{(\phi_k^i \pm \phi_k^j)_{\min}{}^2 + (\psi_k^i \mp \psi_k^j)_{\min}{}^2\}^{1/2}$$

where ϕ_k^j and ψ_k^j are backbone torsion angles of native conformation for the kth residue, ϕ_k^i and ψ_k^i are backbone torsion angles of conformation i for the kth residue, n is number of residues, x equals 180°, and the summation is over all n dihedral angles. The native conformation has the value $\rho = 0$, and ρ increases as the conformations becomes less nativelike. In the case of the enkephalins, there is no one single "native" conformation. The GEM conformation A is an obvious choice, but we also considered, in turn, the other conformations. All backbone torsion angles were included in the calculation except the terminal residues.

Figure 9 gives plots of the average order parameter $\langle\rho\rangle$ in different slabs of potential energy, taking each of the structures A, B, D, and E as the native conformation. Figure 9 also indicates the variability of the individual ρ values within each energy slab (as defined by the error bars plotted at one standard deviation). In the cases where the native conformation is defined as the theoretical minimum energy structure A, $\langle\rho\rangle$ decreases as the energy decreases. In line with the expectation, the internal structure of the conformations gradually becomes more native-like as they move into the native basin. The large variation in the value of ρ in the lowest energy slab indicates that even at the bottom of the funnel there is still a large degree of structural variability. Similarly, $\langle\rho\rangle$ decreases as the energy decreases for other structures B and E. For D, $\langle\rho\rangle$ increases as the energy decreases. For Leu-enkephalin, Figure 10 gives plots of $\langle\rho\rangle$ versus the average potential energy of each slab, taking each of the marked structures as the native conformation. $\langle\rho\rangle$ decreases as energy decreases for A, C, and F. If B is considered the native conformation, $\langle\rho\rangle$ decreases slightly in the second lowest energy bin and then increases as the energy increases. For the structure D, $\langle\rho\rangle$ decreases as energy increases, just as in Met-enkephalin.

The radius of gyration has also been used as an order parameter to plot the free energy landscape in some of the simulation studies.[37] Figure 11A shows the distribution of the

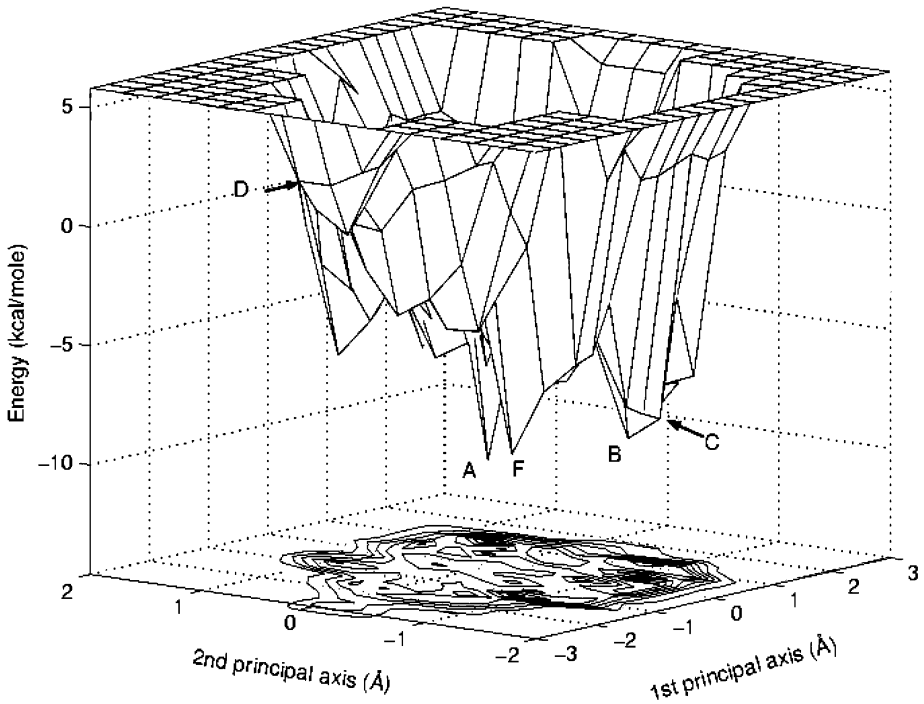

Figure 8. Energy landscape obtained by the minimal energy envelope procedure for Leu-enkephalin.

radius of gyration (R_g) as a function of potential energy for all 427 conformations of Met-enkephalin. The figure shows the expected decrease in radius of gyration with the decrease in energy. However, there are several compact structures, all with R_g less than about 5 Å, but with energies 5−10 kcal/mol greater than the GEM. While the theoretical minimum A has R_g of 4.5 Å, for structure B the value is 4.4 Å, for structure D it is 7.6 Å, and for structure E it is 4.3 Å. Figure 11B shows that, for Leu-enkephalin, the distribution of the radii of gyration is similar to the one for Met-enkephalin. For the representative structure in each group the radius of gyration is as follows: $R_g = 4.4$ Å in group A, 4.8 Å in group B, 4.5 Å in group C, 6.6 Å in group D, and 4.7 Å in group F.

Discussion

Unlike other simulation techniques, the MOLS technique is not a function of time or temperature. The conformational sample we have used to map the energy landscape of the two peptides was generated by a procedure that may be described as sampling a frozen ensemble of all possible conformational states and picking up 1500 structures with the lowest values of the potential function. These 1500 structures are not all unique, and a clustering procedure is used to arrive at the unique set. We observe that the number of low energy states reduces with a decrease in energy, and that compact structures are more favored energetically than extended structures. Even in the second lowest energy bin (Figures 4 and 6), which is only 2.0 kcal/mol higher than the lowest energy bin, the number of mutually different conformations is large. This indicates that there is no deep minimum in the energy landscape, and that the funnel is wide even at the bottom. Figure 12 indicates that fully extended (E) structures (with the distances between $C^{\alpha}_{(1)}$ and $C^{\alpha}_{(4)}$ and between $C^{\alpha}_{(2)}$ and $C^{\alpha}_{(5)}$ >10 Å) occupy the highest energy bins, while fully folded (F) conformations (with these distances <7 Å) populate the lowest energy bins. Partially folded (PF) structures (with the distances between 7 and 10 Å—they are a mixture of 2.2_7 helices or combinations of γ-turns) are

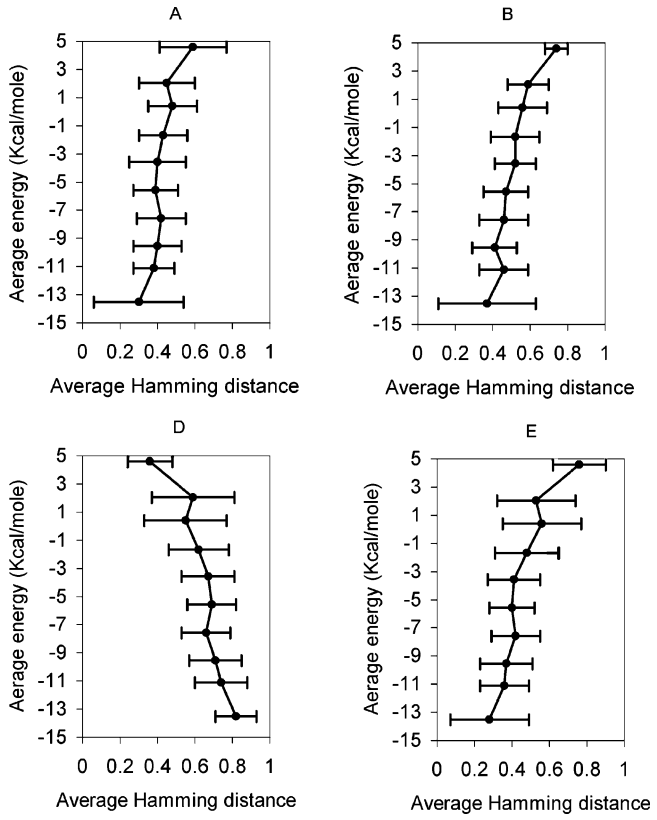

Figure 9. Average ρ (order parameter) versus potential energy for Met-enkephalin. Each point indicates the average ρ calculated over all conformations within a given slab (see Figure 4). The native conformation has a value of $\rho = 0$, and ρ increases as the conformations become less nativelike. The horizontal error bars indicate the structural variability (measured by standard deviation) of individual ρ values within each energy range. The decrease of ρ with decreasing energy indicates a funnel-like landscape, in agreement with Figure 4 for native structure (A). The variation of ρ with energy for other native structures (B), (D), and (E) is also shown.

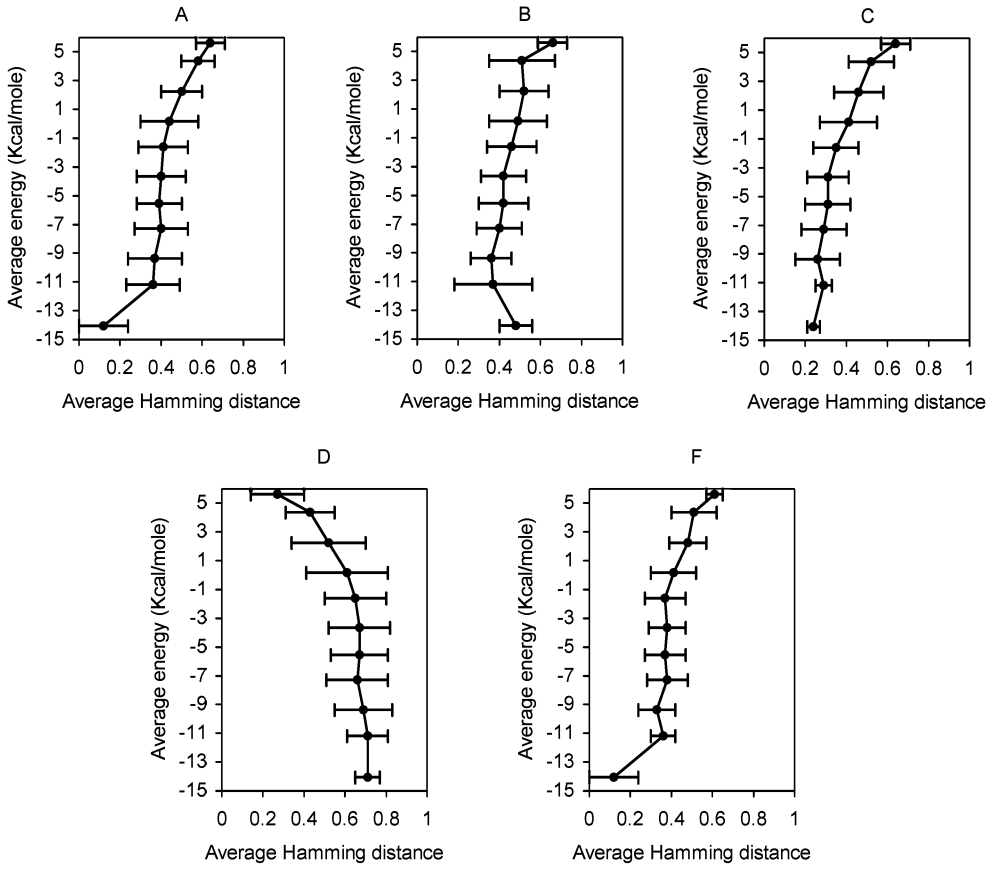

Figure 10. Average ρ versus potential energy values for Leu-enkephalin. The decrease of ρ with decreasing energy indicates a funnel-like landscape, in agreement with Figure 6 for native structure (A). The variation of ρ with energy for other native structures (B), (C), (D), and (F) is also shown.

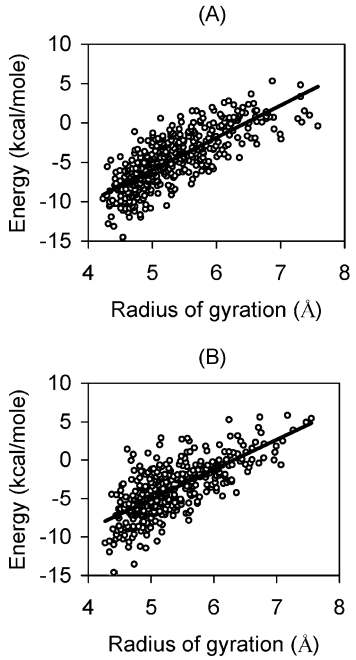

Figure 11. Relative energy as a function of the radius of gyration, R_g. (A) Met-enkephalin; (B) Leu-enkephalin.

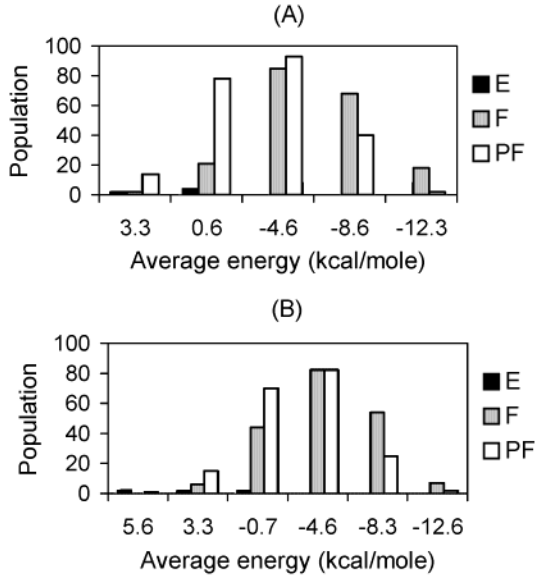

Figure 12. Population of folded (F), partially folded (PF), and extended (E) conformations in each energy bin (with size 4.0 kcal/mol). (A) Met-enkephalin; (B) Leu-enkephalin.

intermediate in energy between the extended and folded states. Also, they have high populations in all the energy bins, with the largest numbers situated in the intermediate energy bins. We therefore conjecture that they could be significant structural intermediates during the folding process. This result is similar to the one obtained by other workers.[13,15,36]

Previous studies on enkephalins have used free energy to draw the landscape, and it is difficult to compare directly our results with theirs.[3,10,15] However, the lowest potential energy basins obtained by our study resembles the lowest free energy states obtained by Evans and Wales.[15] They found two main free energy basins for Met-enkephalin. The lowest basin contains mostly type II′ β-turn structures, similar to the GEM conformation A in the present study. The next lowest free energy region

in their study consists of structures similar to the one marked B here (Figures 4 and 7). We note that the study of Evans and Wales differs from ours not only in the procedure, but also in the model of Met-enkephalin. They use the zwitterionic form of the peptide, whereas we have blocked the terminals with neutral groups.

In molecular simulation studies, the experimental structure (very often the crystal structure) is considered the native state of the molecule. In the case of the enkephalins, this is a puzzle, since experiments have observed ensembles of native conformational families, from extended to folded, while computational searches have often zeroed in on the GEM structure as the native state. Our results show that the lowest energies are populated by many conformations, as observed by other workers as well.[28,32,34,38,39] This multiplicity of favorable structures may be required for the enkephalins to bind multiple receptors[30] (μ, δ, κ, etc).

Acknowledgment. We thank the Council of Scientific and Industrial Research, the Department of Science and Technology, and the Universities Grants Commission, Government of India, for financial support.

References and Notes

(1) Dill, K. A.; Chan, H. S. *Nat. Struct. Biol.* **1997**, *4*, 10−19.
(2) Onuchic, J. N.; Luthey-Schulten, Z.; Wolynes, P. G. *Annu. Rev. Phys. Chem.* **1997**, *48*, 545−600.
(3) Hansmann, U. H. E.; Okamoto, Y.; Onuchic, J. N. *Proteins: Struct., Funct., Genet.* **1999**, *34*, 472−483.
(4) Wales, D. J.; Doye, J. P. K.; Miller, M. A.; Mortenson, P. N.; Walsh, T. A. *Adv. Chem. Phys.* **2000**, *115*, 1−111.
(5) Levy, Y.; Becker, O. M. *J. Chem. Phys.* **2001**, *14*, 993−1009.
(6) Plotkin, S. S.; Onuchic, J. N. *Q. Rev. Biophys.* **2002**, *35*, 111−167; *Q. Rev. Biophys.* **2002**, *35*, 205−286.
(7) Becker, O. M. *J. Mol. Struct.(THEOCHEM)* **1997**, *398*, 507−516.
(8) Bogatyreva, N. S.; Finkelstein, A. V. *Prot. Eng.* **2001**, *14*, 521−523.
(9) Sheinerman, F. B.; Brooks, C. L., III. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 1562−1567.
(10) Klepeis, J. L.; Floudas, C. A. In *Optimization in computational chemistry and molecular biology*; Floudas, C. A., Pardalos, P. M., Eds.; Kluwer Academic Publishers B.V.: New York, 2000; pp 19−46.
(11) Alves, N. A.; Hansmann, H. E. *Int. J. Mod. Phys., C* **2000**, *C11*, 301−308.
(12) Ghosh, A.; Elber, R.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 10394−10398.
(13) Sanbonmatsu, K. Y.; García, A. E. *Proteins: Struct., Funct., Genet.* **2002**, *46*, 225−234.
(14) Arkin, H.; Çelik, T. *Int. J. Mod. Phys., C* **2003**, *14*, 113−120.
(15) Evans, D. A.; Wales, D. J. *J. Chem. Phys.* **2003**, *119*, 9947−9955.
(16) Smith, G. D.; Griffin, J. F. *Science* **1978**, *199*, 1214−1216.
(17) Karle, J. L.; Karle, J.; Mastropaolo, D.; Camerman, A.; Camerman, N. *Acta Crystallogr., Sect. B* **1983**, *39*, 625−637.
(18) Griffin, J. F.; Langs, D. A.; Smith, G. D.; Blundell, T. L.; Tickle, I. J.; Bedarkar, S. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 3272−3276.
(19) Aubry, A.; Birlirakis, N.; Sakarellos-Daitsiotis, M.; Sakarellos, C.; Marraud, M. *Biopolymers* **1989**, *28*, 27−40.
(20) Vengadesan, K.; Gautham, N. *Biophys. J.* **2003**, *84*, 2897−2906.
(21) Vengadesan, K.; Gautham, N. *Biopolymers*, in press.
(22) Fisher, R. A. In *The Design of Experiments*, 7th ed.; Oliver and Boyd: London, 1960; pp 70−92.
(23) Montgomery, D. C. In *Design and Analysis of Experiments*, 5th ed.; John Wiley and Sons: New York, 2000; pp 126−169.
(24) Némethy, G.; Gibson, K. D.; Palmer, K. A.; Yoon, C. N.; Paterlini, G.; Zagari, A.; Rumsey, S.; Scheraga, H. A. *J. Phys. Chem.* **1992**, *96*, 6472−6484.
(25) Betancourt, M. R.; Skolnick, J. *J. Comput. Chem.* **2001**, *22*, 339−353.
(26) Kříž, Z.; Carlsen, P. H. J.; Koca, J. *J. Mol. Struct. (THEOCHEM)* **2001**, *540*, 231−250.
(27) Tuffery, P.; Etchebest, C.; Hazout, S. *Protein Eng.* **1997**, *10*, 361−372.
(28) Isogai, Y.; Némethy, G.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1977**, *74*, 414−418.
(29) Lewis, P. N.; Momany, F. A.; Scheraga, H. A. *Biochim. Biophys. Acta* **1973**, *303*, 211−229.
(30) Schiller, P. W. In *The Peptides*; Udenfriend, S., Meienhofer, J., Eds.; Academic Press: Orlando, 1984; Vol. 6, pp 219−268.
(31) Khaled, M. A. In *Opioid Peptides: Medicinal Chemistry*; Rapaka, R. S., Barnett, G., Hawks, R. L., Eds.; NIDA Research Monograph 69; National Institute on Drug Abuse: Rockville, MD, 1986; pp 266−290.
(32) Eisenmenger, F.; Hansmann, U. H. E. *J. Phys. Chem. B* **1997**, *101*, 3304−3310.
(33) Juárez, R. G.; Morales, L. B.; Pérez, P. F. *J. Mol. Struct. (THEOCHEM)* **2001**, *543*, 277−284.
(34) Paine, G. H.; Scheraga, H. A. *Biopolymers* **1987**, *26*, 1125−1162.
(35) Vasquez, M.; Meirovitch, E.; Meirovitch, H. *J. Phys. Chem.* **1994**, *98*, 9380−9382.
(36) Abdali, S.; Jensen, M. Ø.; Bohr, H. *J. Phys.: Condens. Matter* **2003**, *15*, S1853−S1860.
(37) Hansmann, U. H. E. *Eur. Phys. J. B* **1999**, *12*, 607−611.
(38) Perez, J. J.; Villar, H. O.; Loew, G. H. *J. Comput.-Aided Mol. Des.* **1992**, *6*, 175−190.
(39) Abagyan, R.; Argos, P. *J. Mol. Biol.* **1992**, *225*, 519−532.