JCTC Journal of Chemical Theory and Computation

Effects of Hydration on the Conformational Energy Landscape of the Pentapeptide Met-Enkephalin

L. Ramya and N. Gautham*

Centre of Advanced Study in Crystallography and Biophysics, University of Madras, Chennai, 600025, India

Received January 5, 2009

Abstract: We report here a study of the conformational energy landscape of the pentapeptide Met-enkephalin in the presence of explicit solvent (water) molecules. A sample of 1 500 low-energy structures of this molecule was generated using the mutually orthogonal Latin squares (MOLS) technique with the CHARMM22 force field. This technique, developed in our laboratory, allows us to sample the conformational space of a molecule in an unbiased and exhaustive manner. The study shows that inclusion of explicit solvation is important to correctly model the conformational behavior of the molecule. Structures modeled in the presence of water molecules are far more similar to the experimental structures than when the water molecules are excluded. The results also indicate that the pentapeptide Met-enkephalin prefers extended structures in an aqueous environment, as against tightly folded structures in the absence of water. Thus, the biologically relevant structure, when the molecule is not bound to the receptor, is probably the extended structure, as seen in the crystallographic and NMR experiments, rather than the GEM structures calculated by various workers.

Introduction

Enkephalins are endogenous opioid neuropeptides that have morphine-like activity.[1] One of these, Met-enkephalin, is an inhibitory neurotransmitter, and its action results in decreased pain sensitivity. This pentapeptide takes on different conformations when it binds to the three subclasses of opioid receptors[2] termed μ, δ, and κ based on selectivity of different agonists for different physiological responses.[3] It has been observed that a β-bend folded conformation of the neuropeptide binds to the μ receptor, while an extended conformation binds to the δ receptor.[4] The peptide has a flexible conformation, which is strongly influenced by the local environment. There is evidence that the bound conformations are different from the crystal[5] and solution structures.[6]

In such contexts, computational studies on the peptides become important. The enkephalins, and in particular Met-enkephalin, have been subjected to numerous such studies.[7,8] Many theoretical conformational studies of Met-enkephalin have been carried out using ECEPP, AMBER, and CHARMM force fields.[7] The low-energy conformations of Met-enkephalin calculated in the absence of explicit solvent effects have been determined based on the ECEPP/2 and ECEPP/3 force fields.[8] Eisenmenger and Hansmann[8] have studied the variation of the energy landscape of Met-enkephalin by changing the force field from ECEPP/2 to ECEPP/3. When the force field is changed, the low-energy conformation of the peptide changes, as reflected in the energy landscape, but most of the physical results, such as the thermodynamic properties, are not affected. In the calculations using the ECEPP/2 potential,[9] two different characteristic temperatures of the peptide were determined. These were a collapse temperature T_θ, corresponding to coil-globular transition, and a folding temperature T_f, corresponding to one of the many local minima. Zhan and associates[10] studied the conformations of Met-enkephalin using the ECEPP/2 and ECEPP/3 force fields. With the ECEPP/3 force field, they reported a new global minimum, different from the one located by Isogai et al.[11] All the computations described above were carried out without explicitly specifying the solvent water.

Many other studies have been carried out with implicit solvent.[12−16] Klepsis and Floudas[12] calculated the free energy of the peptide, both with and without solvent, using the αBB

* Corresponding author. E-mail: n_gautham@hotmail.com.

10.1021/ct9000087 CCC: $40.75
Published on Web 07/02/2009

global optimization algorithm. Their results indicated that both cases produced large numbers of low-energy local minima. For the consideration of hydration effects, they adopted a model based on the solvent accessible surface area as well as the solvent accessible volume. Significant differences were observed in the two models, and this led to the conclusion that appropriate model selection is essential for computing hydrated peptide structures. Carlacci[13] studied the thermodynamic properties of the peptide in terms of its ionization state and the solvent environment using Monte Carlo simulated annealing. Again the hydration effects were modeled in terms of the solvent accessible surface area.[14] This analysis supported a mechanism of interconversion between the extended conformer and three bent conformers. The relative free energies of the dominant conformers are not very large, particularly in the solvation calculations. The ensemble entropy is similar in magnitude to the experimental entropy of the unfolding peptide. Shen and Freed[15] performed molecular dynamics simulations for Met-enkephalin with both explicit and implicit solvent models. They observed that the two simulations gave similar results. Simulations of Met-enkephalin have also been carried out using explicit and united atom potentials.[16] The results indicated that the two force fields produced significantly different peptide dynamical properties.

In an earlier report from our laboratory, we explored the conformational space of Met-enkephalin in an exhaustive manner using a technique we had developed called the mutually orthogonal Latin squares technique.[17] The ECEPP/3 potential function was used in the calculations,[18] which were performed with no specific model for water and solvent interactions. It was assumed that the parametrization of the force field based on experimental data automatically incorporated solvent effects. The results of these studies suggested the following: (a) The backbone of the peptide exists as a mixed form of both folded and unfolded structures; (b) The fraction of folded structures is low compared to extended conformations; and (c) Some of the low-energy structures correspond to the fully extended conformation, similar to the crystal structure.

In this paper, we present the results of our calculations of the conformational energy landscape of Met-enkephalin in the presence of explicit water molecules. We perform the calculations using the same mean-field method with MOLS sampling, as in the case described above. However, we have now further enhanced the technique to include explicit water molecules and their interactions. In the Methods and Materials section, we describe the technique, along with details of how it has been modified to include explicit water molecules. The results of the computations are then discussed in the context to the other studies that have been carried out on Met-enkephalin and other peptides, both with and without explicit solvent.

Methods and Materials

The mean-field technique (MFT) with MOLS sampling, used to exhaustively explore the conformational space of oligopeptides, has been extensively described elsewhere.[17−19] A detailed explanation of the MOLS method is also available at http://www.unom.ac.in/Gautham_mols.pdf. For completeness, we repeat the main points of the description here.

MFT[20,21] requires the conformational space to be divided into a number of subspaces, each of which may have a number of states. The effective potential, obtained by setting a particular subspace to a particular state, is then obtained by taking a probability-weighted average of the pairwise interactions between that state of that subspace and all the states of all other subspaces. The probabilities, initially set to be the same for all the states of a given subspace, are then recalculated as the Boltzmann function of the effective potentials. These new probabilities are then used in the next cycle to recalculate the effective potentials. The iterative cycles are repeated until the probabilities (and the effective potentials) converge. The states with the highest probabilities after convergence define the optimum in the conformational space. This procedure has been applied, for example, to determine the optimum side-chain conformations of a protein, given the backbone fold.[21] In this case, the conformation of each side chain constitutes a subspace and the various possible conformers for each constitute its states. The effective potential of setting a particular side-chain to a particular conformer is calculated as the probability-weighted sum of the pairwise interactions between that side chain and each of the conformers of each of the other side-chains in the protein. Upon convergence, the most probable conformers define the structures of the side-chains of the protein.

A crucial factor in the success of such a straightforward application of MFT, in this case, is the relative independence of the side chain structures from each other. Each term in the sum adding up to the effective potential depends only on the conformation of the particular pair of side chains and is independent of the other residues. Thus, the interaction of the residue A with B is considered to depend only on the side chain structures of these two residues and to be independent of the others, an assumption which turns out to be largely valid.

When the method is attempted to be carried over to calculations of the protein backbone conformation, however, a similar assumption of independence does not hold. One may consider each of the backbone torsion angles to be a subspace and the values that are possible for it (for example, any value from 0 to 360° in steps of, say, 10°) as the states accessible to it. However, it is clear that, in this case, the effective potential due to setting one of the torsion angles to one of the values cannot simply be calculated as a weighted sum of pairwise interactions. This is because the interactions that may occur due to the settings of a given pair of torsion angles depend also on the settings of all the intermediate torsions. A rigorous calculation of the effective potential would then require the consideration of all settings of all torsion angles, leading to combinatorial explosion.

To avoid this exponential increase in computation time, we use mutually orthogonal Latin squares (MOLS) sampling to identify a small sample of the conformational space, which is then used to compute the effective potential. MOLS sampling is used in the design of experiments in, for example, agriculture or clinical trials.[22] The common feature in these studies is the large, multivariable experimental space. MOLS

A	B	C
B	C	A
C	A	B

Figure 1. A simple Latin square is of order 3 (adapted from Cochran and Cox[24]).

A α	B β	C γ
B γ	C α	A β
C β	A γ	B α

Figure 2. Two mutually orthogonal Latin squares (MOLS) of order 3. The Latin letters A, B, and C are symbols of the first Latin square. The Greek letters α, β, and γ are symbols of the second Latin square, which is orthogonal to the first Latin square (adapted from Fisher[25]).

is used to identify a small representative sample of this space, and the experiments or calculations are carried out only for this sample, rather than for the entire space. A Latin square of order n is an arrangement of n symbols in a $n \times n$ square array in such a way that each symbol occurs exactly once in every row and column.[23] A simple Latin square of order 3 is shown in Figure 1, where the Latin letters (A, B, and C) are used as symbols to construct the square. Two Latin squares of order n are called orthogonal if, when one is superimposed on the other, all n^2 possible combinations of the two sets of n symbols are present, and each symbol of the one square occurs once, and only once, with each symbol of another square. Figure 2 shows a pair of orthogonal Latin squares involving two sets of three elements each. It was constructed by the superposition of two Latin squares, one square consisting of Latin letters (A, B, and C) and the other consisting of Greek letters (α, β, and γ). It may be seen from Figure 2 that each subsquare consists of one Latin letter and one Greek letter, and each combination of Greek and Latin letters appears only once. In other words the symbols are arranged such that all n^2 possible pairs appear in the array. The concept can be extended to N mutually orthogonal Latin squares (MOLS), every pair of which is orthogonal.[23] Thus, in a set of three MOLS (i.e., $n = 3$), the first square would be orthogonal to both the second and third squares, and the second square would also be orthogonal to the third. It has been shown that if n is a prime number or any integer power of a prime, then it is possible to construct a complete set of $n - 1$ mutually orthogonal Latin squares of order n.[26]

A single Latin square of order n allows one to investigate the effect of setting three subspaces (or factors in the language of experimental design), each at n different states (or values) in only n^2 trials, instead of n^3 trials that would be required for a complete and exhaustive sampling. A set of m mutually orthogonal Latin squares of order n allows one to investigate the effect of setting m factors each to different n values, again in only n^2 (instead of n^m) trials. It may be seen from these descriptions, that a distinctive feature of using MOLS to select the sampled points is the occurrence in the sample of every possible pairwise setting of all the states of all subspaces.

In applying this technique to the exploration of conformational space, including the portion mapped out by the backbone torsion angles of a peptide chain, we identify the subspaces with the torsion angles, there being m such angles, and the states as the values that the torsion angles can take, there being n such values accessible to each torsion angle. Out of the total of n^m points available in the conformational space, we select n^2 points using MOLS and calculate the potential energy corresponding to each of these conformations. To calculate the MFT effective potential due to setting a torsion angle to a specific value, we take the Boltzmann-weighted average of the potential energies at all those n points in the MOLS grid in which that value of that torsion angle occurs. We note that, as described above, this ensures the *pairwise sampling* of that value of the torsion angle with every value of every other torsion angle. This procedure may be repeated for all values of all torsion angles, using the same MOLS grid specifying the n^2 conformations. At the end of this set of calculations, we obtain the effective potentials corresponding to every setting of each torsion angles. In the normal MFT procedure, these potentials are used to recalculate the probabilities, which are, in turn, used to recalculate the effective potentials in iterative cycles until convergence.[21] In the present method using MOLS sampling, however, this is not possible since the effective potential is itself evaluated as a Boltzmann-weighted average of the potentials calculated in the MOLS grid. Thus, the effective potentials at the end of one round of calculations described above are used directly as surrogates for the final probabilities. For each torsion angle, the value with the lowest effective potential is considered the most probable one, and the set of most probable values for all the torsion angles defines the optimum conformation of the peptide.

The crucial assumption underlying the MFT procedure, in particular when used with MOLS sampling, is regarding the degree of independence of the variables (the torsion angles). The procedures work best when the variables are completely independent of each other. However, in that case, they are not required, and a few simple one-dimensional searches would quickly identify the optimum. Backbone torsion angles are highly interdependent, and the procedure outlined above may not lead always to a single global optimum. A second assumption implicit in the procedures is that such a single, global optimum exists on the potential energy landscape. In fact, as many studies have shown,[8] most definitions of the force field operating in proteins and peptides lead to several optima (minima) in conformational space, with approximately equal energy values. We have established[19] that the MOLS sampling procedure offers a way of exploring the entire space in an exhaustive manner (at least for peptides up to about 10 residues in size) and

identifying all the optima. As stated above, the construction of one MOLS grid of order n (and the consequent calculation) leads to one optimal or low-energy conformation. Repeating this procedure by using a different set of MOLS leads to another. There are $(n!)^m$ different ways of constructing the MOLS grid[27] and, in principle, the calculations may be repeated as many times to identify all the low-energy conformations. Of course, this does not mean that there are $(n!)^m$ low-energy conformations for a peptide. Most of the conformations generated in subsequent cycles are the same as one generated previously. We, therefore, use a clustering algorithm[28] to cluster the structures together and identify the unique ones. In practice, we have found that for a pentapeptide it is sufficient to generate 1 500 conformations to identify all its 50 or so low-energy conformations.

We have shown[17] that the sampling of the conformational space of Met-enkephalin carried out by this procedure repeated over a few hundred cycles is exhaustive, and all low-energy structures are identified. The library of structures generated in this manner includes the global energy minima of other simulations as well as the structures obtained by experiments. It also includes some well-folded structures that were not earlier identified by any other experimental or computational method.

In the above reports, we had used the ECEPP/3 semiempirical force field, without including any explicit water molecules. Solvent interactions were deemed to have been incorporated in the force field parametrization. In the present paper, we have extended the method to include explicit water molecules in the calculations. This has been done in the following manner. First, explicit water molecules were added to each of the n^2 MOLS-defined conformations using the program VMD.[29] The peptide was solvated with a water box such that there was a layer of water 5 Å in each direction from the ends of the molecule in each of the three directions. Thus the structure, even if fully extended, was completely immersed in the water. Next, the conformational energy of the molecule was calculated using the CHARMM22 potential.[30] The force field included the torsion angle energy and electrostatic, van der Waals, and solvent interactions. Hydrogen bond energies are implicitly included in the CHARMM potential energy function[30] through nonbonded electrostatic and van der Waals interactions. The TIP3P model[31] of water molecules, as available in the CHARMM force field, was used to calculate the solvent interactions. The n^2 potential energies calculated were then used in the rest of the MOLS procedure to obtain the effective energies for each value of every conformational variable and, subsequently, to identify the optimum conformation, now in the presence of water. The optimal conformation obtained was minimized for 1 000 steps by the conjugate gradient method using the NAMD[32] minimization routine, once again with explicit water molecules.

Using this protocol, 1500 structures were generated for the pentapeptide. These 1 500 minimized structures were clustered using a hierarchical clustering method,[28] with an rmsd cutoff of 1.0 Å for the backbone atoms, i.e., two structures with backbone atom rmsd of 1.0 Å or less were considered to be the same structure and only one of them was retained in the list of unique structures. The procedure

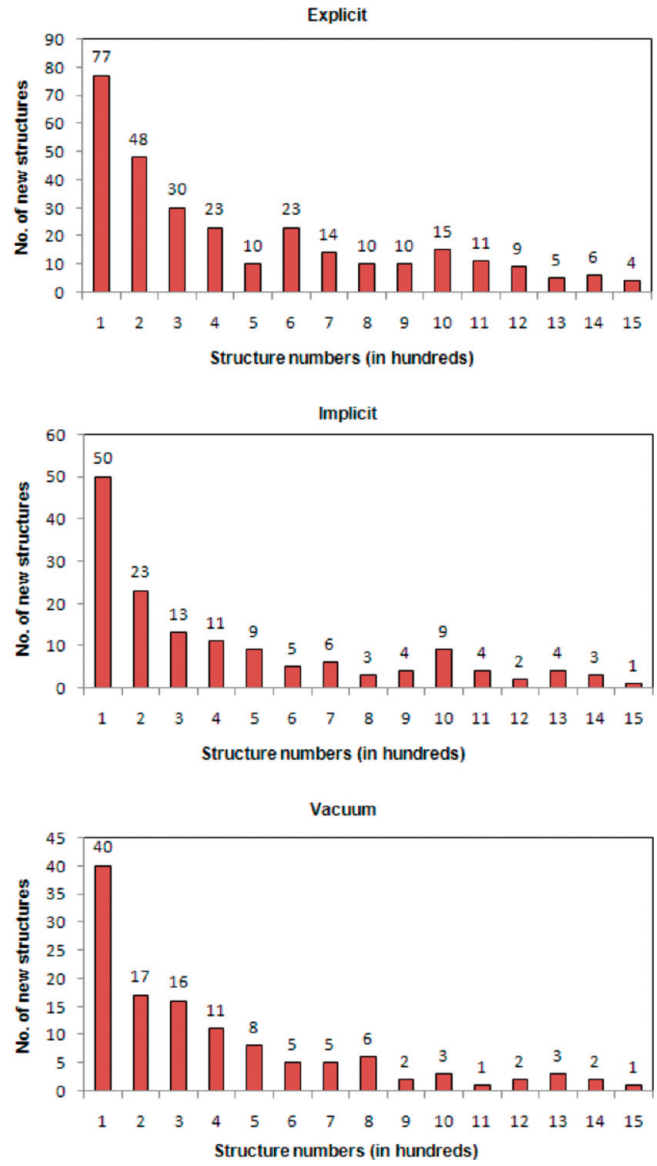

Figure 3. Number of unique structures obtained in every set of 100 structures generated by the MOLS method.

yielded 295 unique structures. Figure 3 shows the number of unique structures obtained in every set of 100 structures generated by the MOLS method. The figure indicates exhaustive sampling of the conformational space of the solvated peptide, since very few new structures are discovered after about 1 200 MOLS structures. The energies used for subsequent comparisons and analyses were those obtained from the NAMD minimization routine, which includes the bond length and angle, dihedral angle, and improper and nonbonded energies. The energies of interaction between the water molecules and the peptide atoms were included in the subsequent analyses but not the intramolecular energies of interaction among the water molecules. This entire procedure was repeated for another set of 1 500 structures; in this case, however, water molecules were not included, thereby yielding a set of structures calculated in vacuum. The set of structures generated by these calculations were clustered using the hierarchical algorithm and resulted in 122 clusters. The computations required 45 min to generate one explicit

solvated structure on an AMD Opteron processor at 2.2 GHz and 7 min on the same platform for each nonsolvated structure.

In order to compare the effects of implicit solvent in the structure prediction, we used the implicit solvent model analytical continuum electrostatic potential[33] (ACE) along with the CHARMM22 force field. (Since NAMD does not support the implicit solvent model, in these calculations we used the program CHARMM in its place, i.e., to calculate the energies and the final minimization step after identifying the low-energy conformations using the MOLS procedure identical to the one above.) As in the case with explicit solvent, 1 500 conformations were generated. These structures were clustered using the hierarchical clustering method with an rmsd cutoff of 1.0 Å for the backbone atoms. This yielded 147 unique structures. The energies used for subsequent comparisons and analyses were those obtained from the CHARMM minimization routine. The computations required 4 min to generate one implicit solvated structure on an AMD Opteron processor at 2.2 GHz.

Results and Discussion

We first discuss the effect of explicit solvation on the global minimum-energy structures, followed by the analyses of the structural motifs in the generated structures. Next, we describe the patterns and distribution of hydrogen bonds. Finally, we describe the effect of the solvation on the energy landscapes.

A. Minimum Energy Structures. Figure 4 compares the minimum-energy structures obtained by the MOLS method with and without solvation. The two structures are quite different from each other, with a rmsd of 3.32 Å on least-squares superposition. In the solvated structure, the side chain of methionine faces the aromatic ring of tyrosine. In the nonsolvated structure, the methionine side chain faces the aromatic ring of phenylalanine. Table 1 gives the backbone rmsd on the superposition of both solvated and nonsolvated lowest-energy structures with the global energy minimum (GEM) structure of Scheraga and co-workers,[11] two NMR structures,[34] and the X-ray crystal structure.[5] The table shows that, first, the nonsolvated structure is closer to the GEM than the solvated one. This is as expected, since the GEM structure was computed in the absence of explicit solvent. Second, when superposed on all three experimental structures, the explicit solvated structure has a smaller rmsd value. Again, this matches with expectations, since the experiments were performed on the solvated molecule.

As noted in our previous computations,[18] of the structures generated, the lowest-energy structure is only very rarely the best-sampled one, i.e., one with the lowest rmsd as compared to either the GEM or the experimental structures. This is true, in the present case too, and Table 2 gives the rmsd and the energy values of the best sampled structures with and without solvent (Figure 5). The rmsd values indicate that the solvated best sampled structures are only marginally closer to the experimental structures and the GEM than are the nonsolvated best sampled structures.

Table 3 gives the results of the comparison of all the MOLS low-energy structures generated, with and without

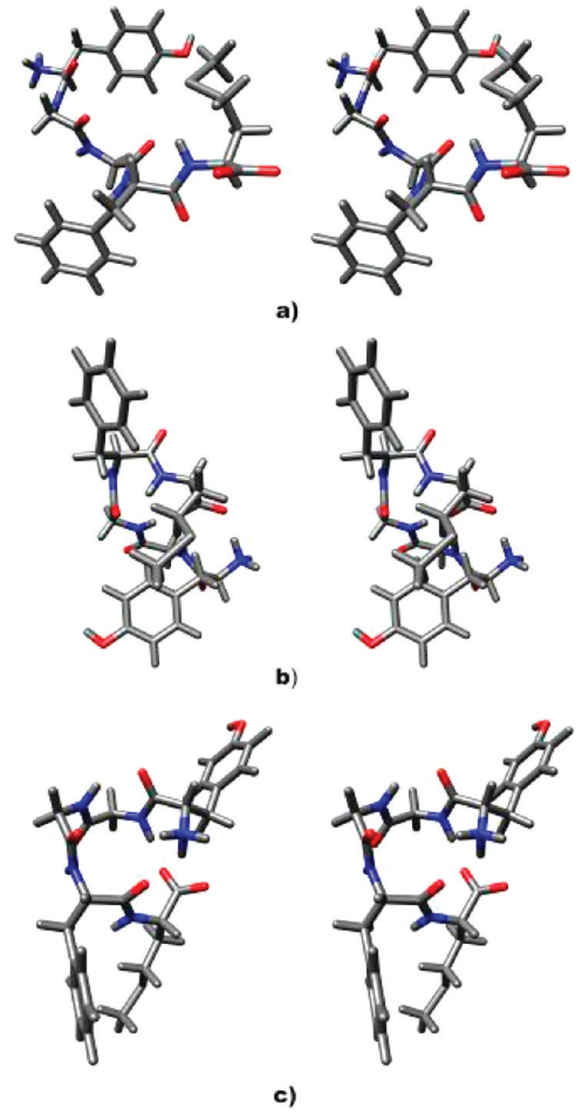

Figure 4. Stereo views of the lowest-energy MOLS (a) explicitly solvated, (b) implicitly solvated, and (c) vacuum structures. The energies of the explicitly solvated, implicitly solvated, and vacuum conformations are −350.6, −80.9, and −50.6 kcal/mol, respectively.

Table 1. Backbone Atoms RMSD (in Å) between MOLS Generated Lowest-Energy Structure and Other Reported Structures[a]

	1PLW	1PLX	FABJIB	GEM
MOLS explicit solvent	2.08	2.13	1.60	2.75
MOLS implicit solvent	2.11	2.22	3.81	1.68
MOLS vacuum	2.59	2.52	3.83	1.37

[a] NMR structures PDB[34] ID: 1PLW; 1PLX and X-ray crystal structure: CSD[5] ID FABJIB; Global minimum energy structure:[11] GEM.

solvent, with the GEM and the experimental structures. It is evident that the solvated structures are, in general, closer to the experimental structures. Likewise, the nonsolvated structures are closer to the GEM. Once again, this matches expectations.

B. Structural Motifs. In order to study the effect of the solvent on the propensity of the molecule to form secondary structural motifs, we used the program PROMOTIF[35] to

Table 2. Best Sampled Structures[a] and Corresponding Energies[b]

	explicit solvent		implicit Solvent		vacuum	
	rmsd (Å)	energy (kcal/mol)	rmsd (Å)	energy (kcal/mol)	rmsd (Å)	energy (kcal/mol)
1PLW	0.73	−246.329	0.90	−70.502	1.01	65.845
1PLX	0.75	−321.723	0.82	−70.502	1.01	65.816
FABJIB	0.36	−308.425	0.33	−70.486	0.39	55.459
GEM	0.53	−218.652	0.69	−74.656	0.57	−39.414

[a] Structures with the lowest rmsd as compared to either the GEM or the experimental structures. Rmsd is calculated for backbone atoms only. [b] NMR structures PDB[34] ID: 1PLW and 1PLX; X-ray crystal structure: CSD[5] ID FABJIB; Global minimum energy structure:[11] GEM.

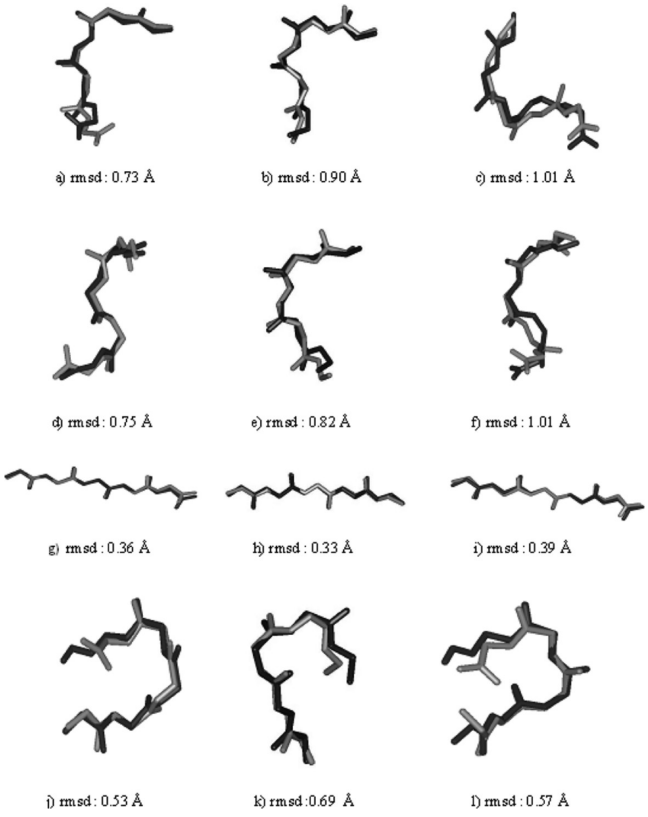

Figure 5. The superposition of the best predicted explicitly solvated, implicitly solvated, and vacuum structures with experimental and GEM structures. Figures (a−c) show superposition of the three structures, respectively, with the NMR structure 1PLW, (d−f) with the NMR structure 1PLX, (g−i) with the crystal structure CSD ID FABJIB, and (j−l) with the GEM structure. Native structures are dark (black), and MOLS structures are light (gray).

identify beta and gamma turns. Folded conformations with hydrogen bonds are termed as turns and the conformations without hydrogen bonds are termed bends.[36] (Both bends and turns are referred to by the general name 'fold'.) A beta turn is defined for four consecutive residues (denoted by i, $i + 1$, $i + 2$, and $i + 3$), if the distance between the C^{α} atom of residue i and the C^{α} atom of residue $i + 3$ is less than 7 Å, and if the central two residues are not helical. A gamma turn is defined for three residues i, $i + 1$, and $i + 2$, if a hydrogen bond exists between residues i and $i + 2$, and the φ and ψ angles of residue $i + 1$ fall within 40° of one of the two sets of values given in Table 4. The analysis shows that the 1 500 explicit solvated structures have a total of 543 beta motifs, of which 56 are beta turns, and the rest are beta bends. The 1 500 nonsolvated structures have 1 410 beta motifs with 82 beta turns and 1328 beta bends. It is apparent from these results that the presence of the water molecules reduces the number of intramolecular interactions, leading to a decrease in the number of secondary structural motifs.

Table 5 shows the distribution of the bends and turns among the various types of beta motifs. There are 32% of the solvated structures that have at least one beta bend or turn, and 75% of the nonsolvated structures have such a feature. Type IV folds constitute the large majority of the folds in both solvated and nonsolvated structures. It has been reported[37] that when the residue at the "$i + 1$" and "$i + 2$" positions is glycine, as in the present case, the peptide has a large propensity to adopt the type IV fold.

In both sets of generated structures, there are many single and double beta folds. There are also several structures with a beta and gamma fold, a beta fold and two gamma folds, two beta folds and a gamma fold, or two beta and gamma folds (Table 5). One of the well-folded solvated structures consists of one 1−4 type IV beta bend, one 2−5 type II′ beta bend, and a single 1−3 inverse gamma turn. This structure, shown in Figure 6, has a low energy of −204.7 kcal/mol. This structure also has one intramolecular hydrogen bond between Tyr^1 (CO) → Gly^3(NH) and six intermolecular hydrogen bonds between the peptide and water molecules. However, it has a backbone rmsd of 1.37 Å with the GEM structure (and 3.45 Å with the crystal structure), indicating that the GEM structure may be different in the presence of solvent. The lowest-energy nonsolvated structure (−50.6 kcal/mol) has two type IV beta bends centered at Gly^2−Gly^3 and Gly^3−Phe^4 and double 1−3 inverse and 2−4 classic gamma turns. This structure, shown in Figure 7, has six hydrogen bonds with a backbone atom rmsd of 1.37 Å with the GEM structure and 3.83 Å with the crystal structure.

C. Hydrogen-Bonding Patterns. *Intramolecular Hydrogen Bonds.* The bonding patterns described by Mitsutake et al.[38] classify 12 possible hydrogen bonds between the backbone atoms in enkephalin. They are N1−O3, O1−N3, N1−O4, O1−N4, N1−O5, O1−N5, N2−O4, O2−N4, N2−O5, O2−N5, N3−O5, and O3−N5. Mitsutake classified Met-enkephalin structures into six classes of similar structures, referred to as C13, C14, C15, C24, C25, and C35. The class is denoted as Cmn where m and n are residues between which the hydrogen bond is formed. For example, a structure that belongs to class C14 would have at least one of the hydrogen bonds N1−O4 or N4−O1. Sometimes both are present giving rise to an antiparallel bonding pattern.

The distribution of intramolecular hydrogen bonds between backbone atoms is tabulated in Table 6. As may be expected, the numbers of such bonds in solvated structures are less than those observed in the nonsolvated structures. Among the solvated structures, the largest numbers of hydrogen bonds occur between residues 3 and 5 and among the nonsolvated structures between residues 1 and 5.

While most of the structures have only a single hydrogen bond, a few have more than one. There are three nonsolvated

Table 3. Total Number of MOLS Structures with Backbone Atoms RMSD Less than 1 Å with Experimental and GEM Structures

	explicit Solvent			implicit Solvent			vacuum		
	no. of structures	avg rmsd (Å)	avg energy (kcal/mol)	no. of structures	avg rmsd (Å)	avg energy (kcal/mol)	no. of structures	avg rmsd (Å)	avg energy (kcal/mol)
1PLW	4	0.88	−285.396	6	0.94	−67.740	nil	–	–
1PLX	3	0.86	−249.127	10	0.89	−69.874	nil	–	–
FABJIB	36	0.86	−277.480	97	0.85	−69.678	37	0.96	59.196
GEM	38	0.89	−211.532	1	0.69	−74.656	109	0.98	−38.103

Table 4. Gamma Turns

turn type	φ $(i+1)$ (deg)	ψ $(i+1)$ (deg)
classic	75.0	−64.0
inverse	−79.0	69.0

structures that have three hydrogen bonds each (Figure 8). In each case, the hydrogen bonds are between the oxygen atom of the terminal carbonyl group and the nitrogen atoms of residues 1, 2, and 3. These three structures have backbone atom rmsd's of 1.34, 1.23, and 1.10 Å, respectively, with the GEM structure and 3.73, 3.66, and 3.58 Å with the crystal structure. These structures have hydrogen bonds between backbone atoms of adjacent residues (Phe^4 (NH) → Met^5 (OT1)).[28]

Apart from the above, there are 10 solvated structures with main chain−side chain hydrogen bonds and 18 nonsolvated structures. Once again, the number is larger for the nonsolvated structures. These results show that, as may be expected, the solvated structures are closer to the experimental structures, which are extended, than to the theoretical GEM

Table 5. Distribution of Turns and Bends

a) Single Beta Turns and Bends

	explicit					implicit					vacuum				
	turns		bends			turns		bends			turns		bends		
type of beta motif	2−3[a]	3−4	2−3	3−4	total	2−3	3−4	2−3	3−4	total	2−3	3−4	2−3	3−4	total
I	1	1	–	1	3	–	–	–	–	–	–	–	–	2	2
II	2	–	2	–	4	1	–	–	–	1	–	–	1	–	1
VIII	1	–	1	7	9	–	–	1	2	3	–	–	2	9	11
I′	–	1	–	–	1	–	–	–	–	–	–	–	–	2	2
II′	2	4	3	1	10	1	–	–	–	1	3	4	3	4	14
IV	15	29	135	337	516	8	5	44	82	139	27	48	433	872	1 380
total	21	35	141	346	543	10	5	45	84	144	30	52	439	889	1 410

b) Single Gamma Turns

	explicit				implicit				vacuum			
type	1−3[b]	2−4	3−5	total	1−3	2−4	3−5	total	1−3	2−4	3−5	total
inverse	74	57	51	182	54	134	86	274	56	35	29	120
classic	63	97	24	184	72	99	25	196	57	186	61	304
total	137	154	75	366	126	233	111	470	113	221	90	424

c) Single and Double Folded Structures

	explicit			implicit				vacuum		
type	single fold	double fold	total	single fold	double fold	triple fold	total	single fold	double fold	total
beta	427	58	485	136	4	–	140	834	288	1 122
gamma	298	34	332	381	40	3	424	340	42	382
total	725	92	817	517	44	3	564	1 174	330	1 504

d) Both Beta Folds and Gamma Turns

	no. of structures		
type	explicit	implicit	vacuum
single β-fold classic γ turn	77	39	179
single β-fold inverse γ turn	38	23	58
single β-fold double γ turns	10	13	18
single β-fold triple γ turns	–	2	–
double β-folds classic γ turn	13	2	82
double β-folds inverse γ turn	2	–	18
double β-folds double γ turns	7	–	24

[a] Residue number of central two residues of turn or bend. [b] Residue number of end residues of turn.

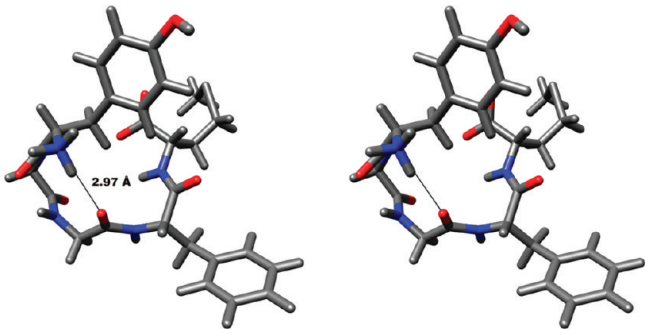

Figure 6. Stereo view of the explicitly solvated structure with two beta bends and an inverse gamma turn. This structure has a backbone rmsd of 1.37 Å with the GEM and 3.45 Å with the crystal structures.

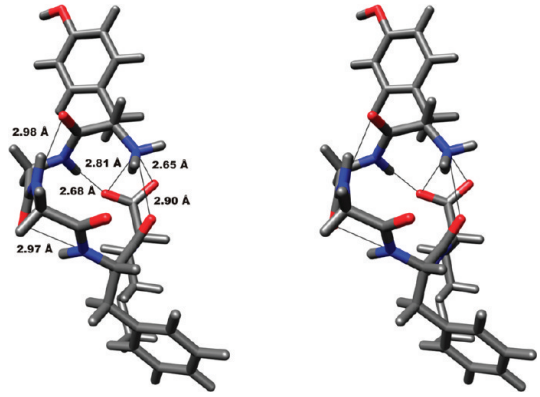

Figure 7. Stereo view of the lowest-energy vacuum structure with two beta bends and two gamma turns.

Table 6. Intramolecular Backbone Hydrogen Bond

type	explicit	implicit	vacuum
C13	2	1	27
C14	15	8	66
C15	59	11	275
C24	2	2	12
C25	61	11	181
C35	63	10	225
total	202	43	786

structure, which is folded. The situation is the reverse for the nonsolvated structures.

Peptide−water Hydrogen-Bonding Pattern. The number of water molecules bound to the peptide range from 5 to 15, with an average of eight waters in each, of the 1 500 solvated structures generated. In general, the extended structures, with lesser number of secondary structure motifs and intramolecular hydrogen bonds, have a greater number of hydrogen-bonded contacts with the water. Figure 9 shows an example. The folded structures, with a larger number of intramolecular hydrogen bonds, make fewer hydrogen-bonded contacts with water (Figure 10). Of the 1 500 generated structures, the one with the lowest energy has 12 hydrogen bonds between the peptide and the solvent water. The structure closest to the crystal structure has 10 peptide−water hydrogen bonds. It is clear that the intramolecular hydrogen bonds in the structures are destroyed by water, which in turn makes it closer to the experimental extended structures.

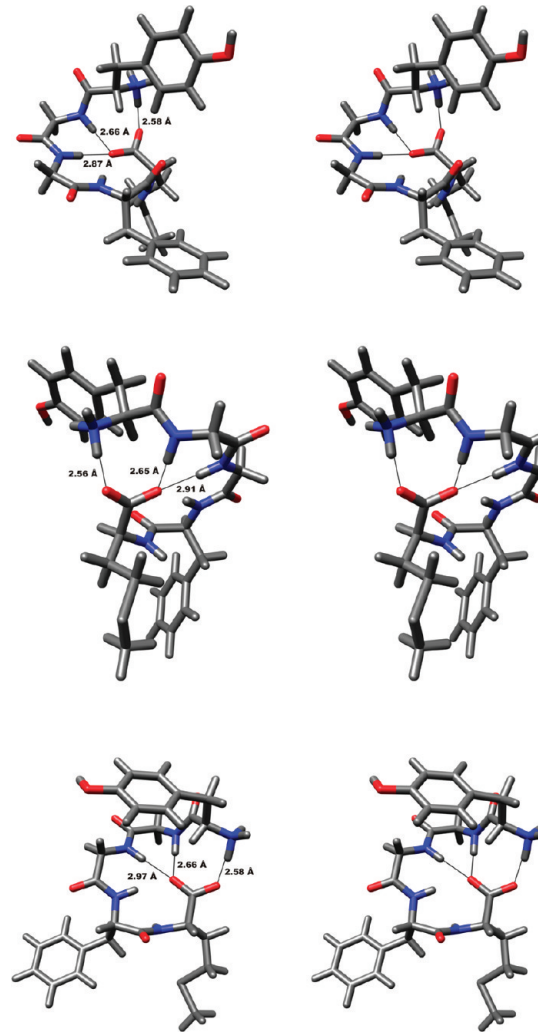

Figure 8. Stereo view of three vacuum structures with three backbone hydrogen bonds formed between residues 1−5, 2−5, and 3−5.

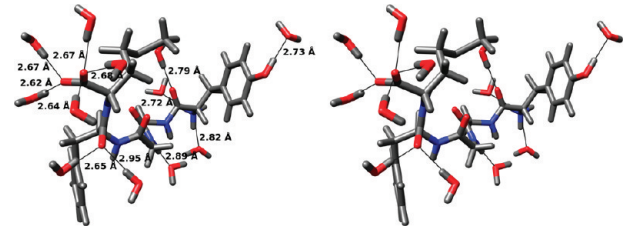

Figure 9. Stereo view of a structure without any intramolecular hydrogen bonds and with a large number of water molecules bonded to the peptide.

D. Energy Landscapes. The potential energy landscape of a peptide may be characterized by the set of locally stable conformations.[39] Since it is impossible to chart molecular energy landscapes in the full 3N-6 dimensional space, a smaller number of dimensions (usually two) is considered sufficient to capture the essential information required for energy landscape mapping, provided these dimensions are chosen carefully. The multidimensional conformational space can be reduced by principal component analysis (PCA)[40] or, more appropriately in the present context, the so-called principal coordinate analysis (PCoorA).[39] In the present case, all the 1 500 structures generated in each set, solvated and

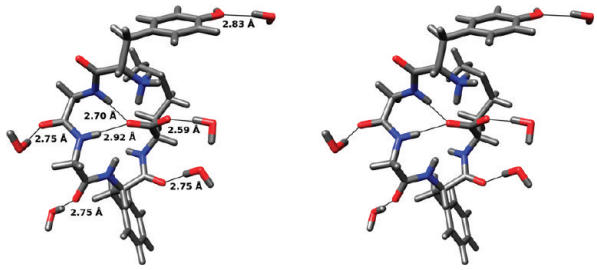

Figure 10. Stereo view of a structure with two intramolecular hydrogen bonds and few peptide−water hydrogen bonds.

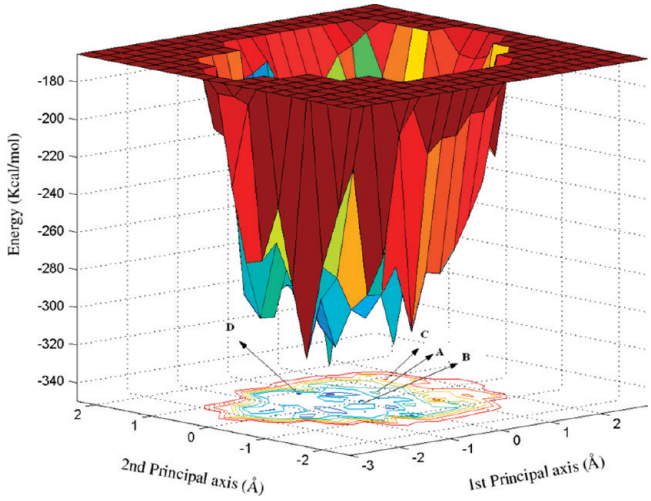

Figure 11. Energy landscape obtained for Met-enkephalin in the presence of explicit water. The two principal axes indicate conformational similarity, and the vertical axis indicates the relative energy.

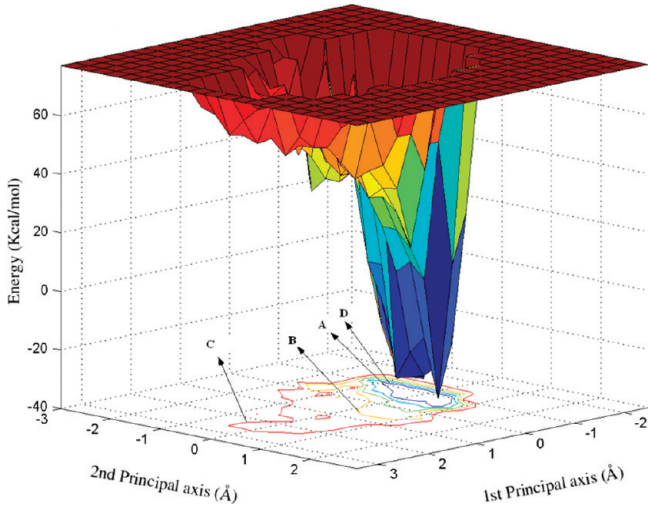

Figure 12. Energy landscape obtained for Met-enkephalin in the absence of water.

nonsolvated, were used to construct the respective energy landscapes. After the PCoorA, the first two principal coordinates accounted for 36.2% of the variance in the solvated set of structures and 49.1% of the variance in the nonsolvated set of structures. The energy landscapes for these three sets are shown in Figures 11 and 12. Clearly, solvation increases the width of the 'folding funnel'[41] as well as the ruggedness at the bottom of the funnel. There are many more

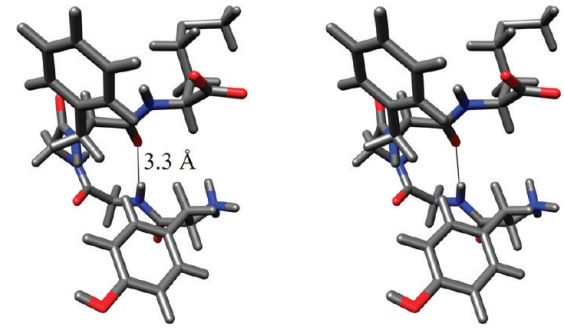

Figure 13. Stereo view of the implicitly solvated structure with two beta bends and a classic gamma turn. This structure has a backbone rmsd of 1.78 Å with the GEM and 3.46 Å with the crystal structures.

minima upon solvation than when the peptide is not solvated, though the energy differences between these minima are small. The respective energy minima are marked on the *xy* plane in both figures (the arrows labeled 'A'), as are the points representing the experimental structures (arrows labeled 'B' for the NMR structure and 'C' for the X-ray crystal structure), and the GEM (arrow labeled 'D'). In the energy landscape of the solvated structures the minimum-energy structure is closer to the experimental structures than to the GEM. It is vice versa in the landscape of nonsolvated structures.

E. Effects of the Implicit Solvent Model. The lowest-energy structure with implicit solvent has an energyof −80.9 kcal/mol. This structure is different from both the one generated with an explicit solvent (backbone rmsd of 2.94 Å) and the one generated in vacuum (1.99 Å). Table 1 shows that the lowest-energy implicit solvent structure is close to the GEM rather than the experimental structures, though among the 1 500 low-energy structures generated, there is a greater number closer to the experimental structures (Table 3).

Among these 1 500 structures, there are 144 beta motifs, of which 15 are beta turns and the rest are beta bends. There are 9% of the structures with an implicit solvent have at least one beta bend or turn. One of the well-folded structures consists of one 1−4 type IV beta bend, one 2−5 type IV beta bend, and one 2−4 classic gamma turn. This structure, shown in Figure 13, has an energy of −71.7 kcal/mol. It has a backbone rmsd of 1.78 Å with the GEM structure and 3.46 Å with the crystal structure.

The number of hydrogen bonds observed in these structures is even less than that observed in structures with an explicit solvent. The largest number of hydrogen bonds is between residues 1 and 5 or 2 and 5. There are three structures that have three hydrogen bonds each (Figure 14). In each case, the hydrogen bonds are between the oxygen atom of the terminal carbonyl group and the nitrogen atoms of residues 1, 2, and 3. These three structures have backbone rmsd of 1.09, 1.68, and 1.30 Å, respectively, with the GEM structure and 3.83, 3.81, and 3.77 Å with the crystal structure. Apart from the above, there are three structures with main chain−side chain hydrogen bonds.

All the 1 500 structures generated with an implicit solvent were used to construct the energy landscape. After the PCoorA procedure, the first two principal coordinates ac-

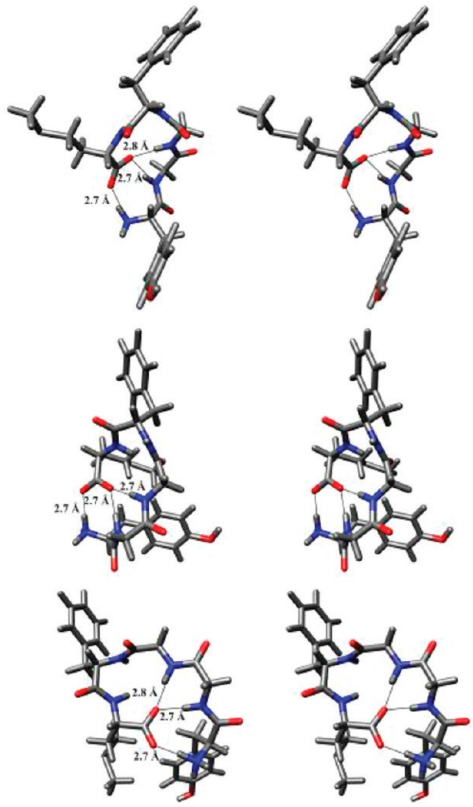

Figure 14. Stereo view of three implicitly solvated structures with three backbone hydrogen bonds formed between residues 1−5, 2−5, and 3−5.

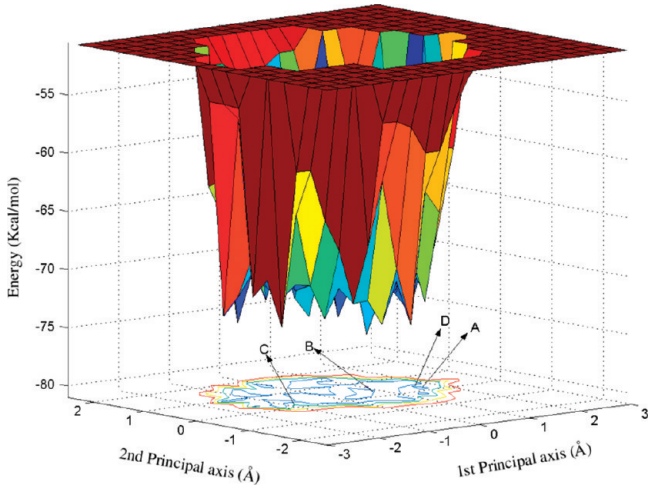

Figure 15. Energy landscape obtained for Met-enkephalin in the presence of implicit water.

counted for 27.4% of the conformational variance in the structures. Figure 15 shows the energy landscape drawn along these two principal coordinates. It may be seen that the effect of the implicit solvent are, in general, the same as that of an explicit solvent. Again, there are many more minima upon solvation than when the peptide is not solvated, though the energy differences between these minima are small.

Conclusion

We have used the MOLS algorithm to study the effect of solvation on the conformational energy landscape of Met-enkephalin. The MOLS algorithm is particularly well suited for this study since it performs an exhaustive and unbiased sampling of the conformational space of the peptide to locate all the low-energy conformers. The study shows that inclusion of solvation is important to correctly model the conformational behavior of the molecule. Structures modeled in the presence of water molecules are far more similar to the experimental structures than when the water molecules are excluded.

The calculations also indicate that the pentapeptide Met-enkephalin prefers extended structures in an aqueous environment, as against tightly folded structures in the absence of water. Thus, the biologically relevant structure may be the extended structure as seen in the crystallographic and NMR experiments,[5,34] rather than the GEM structures calculated by various workers.[11] The calculations are performed on the unbound molecule, whereas, as stated earlier in the Introduction, various workers[4] have described a folded conformation for the molecule bound to the receptor protein. One may speculate that Met-enkephalin exists as an extended molecule in solution but assumes a folded form on binding to the receptor.

Acknowledgment. We thank the Department of Science and Technology, Government of India and UGC, Government of India under Centre of Advanced Study program for financial support.

References

(1) Hughes, J.; Smith, T. W.; Kosterlitz, H. W.; Fothergill, L. A.; Morgan, B. A.; Morris, H. R. Identification of two related pentapeptides from the brain with potent opiate agonist activity. *Nature (London).* **1975**, *258*, 577–579.

(2) Schiller, P. W. Conformational analysis of enkephalin and conformation-activity relationships. In *The Peptides: Analysis, Synthesis, Biology*; Udenfriend, S., Meinhofer, J., Eds.; Academic Press: Orlando, FL, 1984; Vol. 6, pp 219−268.

(3) Atweh, S. F.; Kuhar, M. J. Distribution and physiological significance of opioid receptors in the brain. *Br. Med. Bull.* **1983**, *39*, 47–52.

(4) Ishida, T.; Kenmotsu, M.; Mino, Y.; Inoue, M.; Fujiwara, T.; Tomita, K.-I.; Kimura, T.; Sakakibara, S. X-ray diffraction studies of enkephalins. Crystal structure of [(4′-bromo) Phe 4, Leu 5] enkephalin. *Biochem. J.* **1984**, *218*, 677–689.

(5) Griffin, J. F.; Langs, D. A.; Smith, G. D.; Blundell, T. L.; Tickle, I. J.; Bedarkar, S. The crystal structures of [Met^5] enkephalin and a third form of [Leu^5] enkephalin: Observations of a novel pleated β-sheet. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 3272–3276.

(6) Graham, W. H.; Carter, E. S., II.; Hicks, 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.

(7) Montcalm, T.; Cui, W.; Zhao, H.; 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.

(8) 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**, *10*, 3304–3310.

(9) Hansmann, U. H. E.; Masuya, M.; Kamoto, Y. Characteristic temperatures of folding of a small peptide. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 10652–10656.

(10) Zhan, L.; Chen, J. Z. Y.; Liu, W. Conformational study of Met-enkephalin based on the ECEPP force fields. *Biophys. J.* **2006**, *91*, 2399–2404.

(11) Isogai, Y.; Nemethy, G.; Scheraga, H.A. Enkephalin: conformational analysis by means of empirical energy calculations. *Proc. Natl. Acad. Sci. U.S.A.* **1977**, *74*, 414−418.

(12) Klepeis, J. L.; Floudas, C. A. Free-energy calculations for peptides via deterministic global optimization. *J. Chem. Phys.* **1999**, *110*, 7491–7512.

(13) Carlacci, L. Conformational analysis of [Met^5]-enkephalin: solvation and ionization considerations. *J. Comput.−Aided Mol. Design* **1998**, *12*, 195–213.

(14) Ooi, T.; Oobatake, M.; Nemethy, G.; Scheraga, H. A. Accessible surface areas as a measure of the thermodynamic parameters of hydration of peptides. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 3086–3090.

(15) Shen, M.-Y.; Freed, K. F. Long time dynamics of Met-enkephalin: comparison of explicit and implicit solvent models. *Biophys. J.* **2002**, *82*, 1791–1808.

(16) Zaman, M. H.; Shen, M.-Y.; Berry, R. S.; Freed, K. F. Computer simulation of Met-enkephalin using explicit atom and united atom potentials: similarities, differences, and suggestions for improvement. *J. Phys. Chem. B* **2003**, *107*, 1685–1691.

(17) Vengadesan, K.; Gautham, N. Enhanced sampling of the molecular potential energy surface using mutually orthogonal Latin squares: Application to peptide structures. *Biophys. J.* **2003**, *84*, 2897–2906.

(18) Vengadesan, K.; Gautham, N. Conformational studies on enkephalins using the MOLS technique. *Biopolymers* **2004**, *74*, 476–494.

(19) Vengadesan, K.; Gautham, N. Energy landscape of Met-enkephalin and Leu-enkephalin drawn using mutually orthogonal Latin squares sampling. *J. Phys. Chem. B* **2004**, *108*, 11196–11205.

(20) Olszewski, K. A.; Piela, L.; Scheraga, H. A. Mean field theory as a tool for intramolecular conformational optimization. 1. Tests on terminally-blocked alanine and Met-enkephalin. *J. Phys. Chem.* **1992**, *96*, 4672–4676.

(21) Koehl, P.; Delarue, M. Application of a self-consistent mean field theory to predict protein side-chains conformation and estimate their conformational entropy. *J. Mol. Biol.* **1994**, *239*, 249–275.

(22) Finney, D. J. Randomized blocks and Latin squares. In *Experimental Design and its Statistical Basis*; Cambridge University Press: London, U.K., 1955; pp 45−67.

(23) Ito, K. Latin squares. In *Encyclopedic Dictionary of Mathematics*; MIT Press: Cambridge, MA, 1987; Vol. 2, pp 891−892.

(24) Cochran, W. G.; Cox, G. M. *Experimental Designs*, 2nd ed.; John Wiley and Sons: New York, 1957; pp 117−133.

(25) Fisher, R. A. *The Design of Experiments*, 7th ed.; Oliver and Boyd: London, 1960; pp 70−92.

(26) Ryser, H. J. *Combinatorial Mathematics*, Mathematical Association of America: Washington, DC, 1963; pp 79−84.

(27) Liu, C. L. *Introduction to Combinatorial Mathematics*, McGraw-Hill Book Company: New York, 1968; pp 1−22.

(28) Kriz, Z.; Carlsen, P. H. J.; Koca, J. Conformational features of linear and cyclic enkephalins. A computational study. *J. Mol. Struct. (THEOCHEM)* **2001**, *540*, 231–250.

(29) Humphrey, W.; Dalke, A.; Schulten, K. VMD - Visual molecular dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38.

(30) MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E., III; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

(31) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935.

(32) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781–1802.

(33) Schaefer, M.; Karplus, M. A comprehensive analytical treatment of continuum electrostatics. *J. Phys. Chem.* **1996**, *100*, 1578–1599.

(34) Marcotte, I.; Separovic, F.; Auger, M.; Gagne, S. M. A multidimensional 1H NMR investigation of the conformation of methionine-enkephalin in fast-tumbling. *Biophys. J.* **2004**, *86*, 1587–1600.

(35) Hutchinson, E. G.; Thornton, J. M. PROMOTIF - A program to identify and analyze structural motifs in proteins. *Protein Sci.* **1996**, *5*, 212–220.

(36) Van der Spoel, D.; Berendsen, H. J. C. Molecular dynamics simulations of Leu-enkephalin in water and DMSO. *Biophys. J.* **1997**, *72*, 2032–2041.

(37) Lewis, P. N.; Momany, F. A.; Scheraga, H. A. Chain reversals in proteins. *Biochim. Biophys. Acta* **1973**, *303*, 211–229.

(38) Mistsutake, A.; Hansmann, U. H. E.; Okamoto, Y. Temperature dependence of distributions of conformations of a small peptide. *J. Mol. Graphics Modell.* **1998**, *16*, 226–238.

(39) Levy, Y.; Becker, O. M. Energy landscapes of conformationally constrained peptides. *J. Chem. Phys.* **2001**, *114*, 993–1009.

(40) Becker, O. M.; Levy, Y.; Ravitz, O. Flexibility, conformation spaces and bioactivity. *J. Phys. Chem. B* **2000**, *104*, 2123–2135.

(41) Garcia, A. E.; Sanbonmatsu, K. Y. Exploring the energy landscape of a β hairpin in explicit solvent. *Proteins: Struct., Funct., Genet.* **2001**, *42*, 345–354.

CT9000087