First-Principle Determination of Peptide Conformations in Solvents: Combination of Monte Carlo Simulated Annealing and RISM Theory

Masahiro Kinoshita,*,† Yuko Okamoto,‡ and Fumio Hirata‡

Contribution from the Advanced Energy Utilization Division, Institute of Advanced Energy, Kyoto University, Uji, Kyoto 611, Japan, and Department of Theoretical Studies, Institute for Molecular Science, Okazaki, Aichi 444, Japan

Received June 23, 1997. Revised Manuscript Received December 5, 1997

Abstract: This paper contributes to development of a microscopic approach to predicting stable conformations of proteins in solvent. We report results of the first attempt to combine Monte Carlo simulated annealing, a powerful conformational sampling technique, and the reference interaction site model (RISM) theory, a statistical-mechanical treatment for molecular fluids. In solvent the key function is the total energy defined as the sum of the conformational energy and the solvation free energy, and the RISM theory is employed to calculate the latter. Starting from an initial conformation given, our computer program samples many conformations and then finds the conformation with the minimum total energy. Met-enkephalin in the two different solvents, a model water and a simple, repulsive-potential system, are considered. In water the solvation free energy varies greatly from conformation to conformation, while in the simple solvent it remains almost unchanged against conformational changes. In water most of the conformations with larger solvation free energies are strongly rejected, and the number of probable conformations is drastically reduced, which is suggestive that Met-enkephalin is forced to take conformations favored by water far more rapidly than in gas phase and in the simple solvent. The set of stable conformations obtained in water are quite different from those in gas phase and the simple solvent: they are characterized by almost fully extended backbone structure with large fluctuations in side-chain structure, which are in qualitatively good agreement with those determined by the recent nuclear magnetic resonance (NMR) experiments.

Introduction

The first-principle prediction of tertiary structures (conformations) of proteins in solvent from their primary structures is one of the most challenging problems in molecular science. There are two major difficulties to be overcome in solving this problem. First, the number of possible conformations is astronomically large. Second, effects due to the solvent can be substantial and need to be taken into account in an explicit manner. To overcome the first difficulty, powerful conformational sampling methods were developed. The most commonly used one may be simulated annealing.[1] Generalized-ensemble algorithms, among which multicanonical approach[2] is well-known, are also effective (see ref 3 for one of the latest such methods). The usefulness of these methods was demonstrated for problems of polypeptide conformation prediction in gas phase.[4−8] As for the second difficulty, we propose the reference interaction site model (RISM) theory[9−12] with the hypernetted-chain (HNC) approximation as a reliable tool. In our earlier work[13,14] a robust algorithm was developed for solving the full RISM equations (in the full RISM theory, the superposition approximation[15] in which the entire free energy of a peptide is expressed as a sum of the potential of mean forces between pairs of atoms, is not employed). The algorithm was applied to analyses of the solvation structure and conformational stability of Met-enkephalin (Tyr-Gly-Gly-Phe-Met) in the extended simple point charge (SPC/E) model[16] water. Four different conformations of Met-enkephalin were considered, and the solvation structure was analyzed at an atomic level. It was shown that our algorithm is orders of magnitude faster than the conventional one and the RISM theory is a promising tool for calculating the relative values of solvation free energies among different peptide conformations.[13,14]

† Kyoto University.
‡ Institute for Molecular Science.

(1) Kirkpatrick, S.; Gelatt, C. D., Jr.; Vecchi, M. P. *Science* **1983**, *220*, 671−680.
(2) (a) Berg, B. A.; Neuhaus, T. *Phys. Lett.* **1991**, *B267*, 249−253. (b) Berg, B. A.; Neuhaus, T. *Phys. Rev. Lett.* **1992**, *68*, 9−12.
(3) Hansmann, U. H. E.; Okamoto Y. *Phys. Rev. E* **1997**, *56*, 2228−2233.
(4) Wilson, S. R.; Cui, W.; Moskowitz, J. W.; Schmidt, K. E. *Tetrahedron Lett.* **1988**, *29*, 4373−4376.
(5) Kawai, H.; Kikuchi, T.; Okamoto, Y. *Protein Eng.* **1989**, *3*, 85−94.
(6) Hansmann, U. H. E.; Okamoto, Y. *J. Comput. Chem.* **1993**, *14*, 1333−1338.
(7) Hao, M. H.; Scheraga, H. A. *J. Phys. Chem.* **1994**, *98*, 4940−4948.
(8) Hansmann, U. H. E.; Okamoto Y. *J. Comput. Chem.* **1997**, *18*, 920−933.
(9) Chandler, D.; Andersen, H. C. *J. Chem. Phys.* **1972**, *57*, 1930−1937.
(10) (a) Hirata, F.; Rossky, P. J. *Chem. Phys. Lett.* **1981**, *83*, 329−334. (b) Hirata, F.; Levy, R. M. *Int. J. Quantum Chem.* **1988**, *15*, 179−190.
(11) (a) Perkyns, J.; Pettitt, B. M. *Chem. Phys. Lett.* **1992**, *190*, 626−630. (b) Perkyns, J.; Pettitt, B. M. *J. Chem. Phys.* **1992**, *97*, 7656−7666.
(12) Kinoshita, M.; Hirata, F. *J. Chem. Phys.* **1997**, *106*, 5202−5215.
(13) Kinoshita, M.; Okamoto, Y.; Hirata, F. *J. Comput. Chem.* **1997**, *18*, 1320−1326.
(14) Kinoshita, M.; Okamoto, Y.; Hirata, F. *J. Chem. Phys.* **1997**, *107*, 1586−1599.
(15) (a) Pettitt, B. M.; Karplus, M. *Chem. Phys. Lett.* **1985**, *121*, 194−201. (b) Ramé, G. L.; Lau, W. F.; Pettitt, B. M. *Int. J. Peptide Protein Res.* **1990**, *35*, 315−327.
(16) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.* **1987**, *91*, 6269−6271.

S0002-7863(97)02048-9 CCC: $15.00
Published on Web 02/06/1998

Behaviors of a solute molecule in solvent can be quite different from those in gas phase. The solute molecule itself tends to take conformations whose conformational energy is as low as possible. The solvent, on the other hand, prevents the solute molecule from taking conformations which largely perturb the solvent structure, i.e., those with high solvation free energies: the solvent tries to force the solute molecule to take conformations with the lowest solvation free energy. The conformation of the solute molecule in the solvent is determined from the competition of these two factors. Hence, the total energy, the sum of the conformational energy and the solvation free energy, is the key function in analyses of the conformational stability in the solvent. Our ultimate goal is to combine the fast solution algorithm for the RISM theory with the powerful conformational sampling methods mentioned above so that the conformational stability of a protein in aqueous electrolyte solutions can be analyzed under a variety of environmental conditions. As an essential step in this direction, Met-enkephalin (the number of the atomic sites is 75) in pure solvent is chosen in the present article, and Monte Carlo (MC) simulated annealing and the RISM theory are combined for the first time. We consider two different solvents: the SPC/E water and a simple, repulsive-potential system. The latter can be a simple model for nonpolar solvents. Starting from an initial conformation given, our computer program samples many conformations in accordance with the simulated annealing technique and then finds the lowest-energy conformation (i.e., the conformation with the minimum total energy). Several different conformations are tested as the initial ones. The results obtained in the water case are compared with those in the simple-solvent case and with the observations in the recent nuclear magnetic resonance (NMR) experiments.[17] The uniqueness of water as a solvent is emphasized.

Methodology

We assume that the solute molecule (i.e., a peptide in solvent) has m atomic sites. The solvation free energy for the solute molecule $\Delta\mu_s$ is calculated from[13,14,18]

$$\frac{\Delta\mu_s}{k_B T} = 4\pi \int_0^\infty F(r)\, \mathrm{d}r \tag{1}$$

$$F(r) = \sum_{A=1}^{m} \sum_{B=H,O} \rho_B r^2 \left[\frac{\{h_{AB}(r)\}^2}{2} - c_{AB}(r) - \frac{h_{AB}(r) c_{AB}(r)}{2} \right] \tag{2}$$

$$\rho_H = 2\rho_O \tag{3}$$

where k_B, T, and ρ are the Boltzmann constant, absolute temperature, and number density, respectively. The subscripts "A" and "B" denote the atomic sites in the peptide and those in water, respectively. The site−site intermolecular correlation functions, $h_{AB}(r)$ and $c_{AB}(r)$, are calculated by solving the RISM-HNC equations.[13,14] We note that $\Delta\mu_s$ is dependent on the conformation of the solute molecule.

The model of a water molecule is the SPC/E model.[16] The site−site pair interactions $u_{AB}(r)$ has the form

$$u_{AB}(r) = \frac{q_A q_B}{r} + 4\epsilon_{AB}\left\{ \left(\frac{\sigma_{AB}}{r}\right)^{12} - \left(\frac{\sigma_{AB}}{r}\right)^{6} \right\} \tag{4}$$

$$A = 1, ..., m; \quad B = H,O$$

where q_A is the partial charge on site A of the solute molecule, and the standard combination rule

$$\begin{cases} \epsilon_{AB} = \sqrt{\epsilon_A \epsilon_B} \\ \sigma_{AB} = \dfrac{\sigma_A + \sigma_B}{2} \end{cases} \tag{5}$$

is employed for calculating the Lennard-Jones potential parameters. The peptide we consider in the present article is Met-enkephalin (m = 75). The potential-energy functions and parameters are adopted from KONF90 (ref 19) which is based on ECEPP/2 (ref 20). The values of q_A and σ_A for Met-enkephalin are given in our previous paper.[14] The dimensionless number density of water $\rho_O d^3$ (d = 0.28 nm) is 0.7317. We also consider a simple, repulsive-potential system as the solvent. The particles of this solvent interact through

$$u(r) = 4\epsilon\left(\frac{\sigma}{r}\right)^{12} \tag{6}$$

where ϵ = 0.156 kcal/mol and σ = 0.28 nm. The dimensionless number density in the bulk is taken to be 0.7317 (the same as the water value). The interaction between the solvent particle and an atomic site of the peptide is also expressed as the form of eq 6 (i.e., only the repulsive part of the Lennard-Jones potential is considered) with the combination rule of eq 5.

The computer program[13] for solving the RISM equations and calculating the solvation free energy has been incorporated in the program of MC simulated annealing,[19] and a combined program has thus been developed. The scheme for MC simulated annealing is the same as that employed in our earlier work[19] except that the conformational energy E_C is replaced by the total energy E_T defined by

$$E_T = E_C + \Delta\mu_S \tag{7}$$

We note that the important quantities are not the absolute values of E_C, $\Delta\mu_S$, and E_T but the relative values among different conformations of the peptide.

Starting from an initial conformation given, the combined program samples many conformations in accordance with the simulated annealing scheme and then finds the conformation with the minimum value of E_T. One MC sweep updates all the torsion angles of the peptide once. The initial and final temperatures for the conformational sampling[1] are set to 500 and 300 K, respectively. The temperature is decreased exponentially with the MC sweeps. The computation is performed on our workstations (IBM RS6000/3CT; 64MB and 128MB). The RISM equations are fully solved to calculate $\Delta\mu_S$ for each conformation sampled. Employment of eq 7 implies that the entire configuration space for the solvent, which is in equilibrium with each conformation of the peptide, is efficiently sampled at each trial step. In this respect our method predominates over the usual computer simulation methods performed for the whole

(17) Graham, W. H.; Carter II, E. S.; Hickes, R. P. *Biopolymers* **1992**, *32*, 1755−1764.

(18) (a) Zichi, D. A.; Rossky, P. J. *J. Chem. Phys.* **1986**, *84*, 1712−1723. (b) Yu, H. A.; Roux, B.; Karplus, M. *J. Chem. Phys.* **1990**, *92*, 5020−5033.

(19) (a) Kawai, H.; Okamoto, Y.; Fukugita, M.; Nakazawa, T.; Kikuchi, T. *Chem. Lett.* **1991**, 213-216. (b) Okamoto, Y.; Fukugita, M.; Nakazawa, T.; Kawai, H. *Protein Eng.* **1991**, *4*, 639−647.

(20) (a) Momany, F. A.; McGuire, R. F.; Burgess, A. W.; Scheraga, H. A. *J. Phys. Chem.* **1975**, *79*, 2361−2381. (b) Némethy, G.; Pottle, M. S.; Scheraga, H. A. *J. Phys. Chem.* **1983**, *87*, 1883−1887. (c)Sippl, M. J.; Némethy, G.; Scheraga, H. A. *J. Phys. Chem.* **1984**, *88*, 6231−6233.

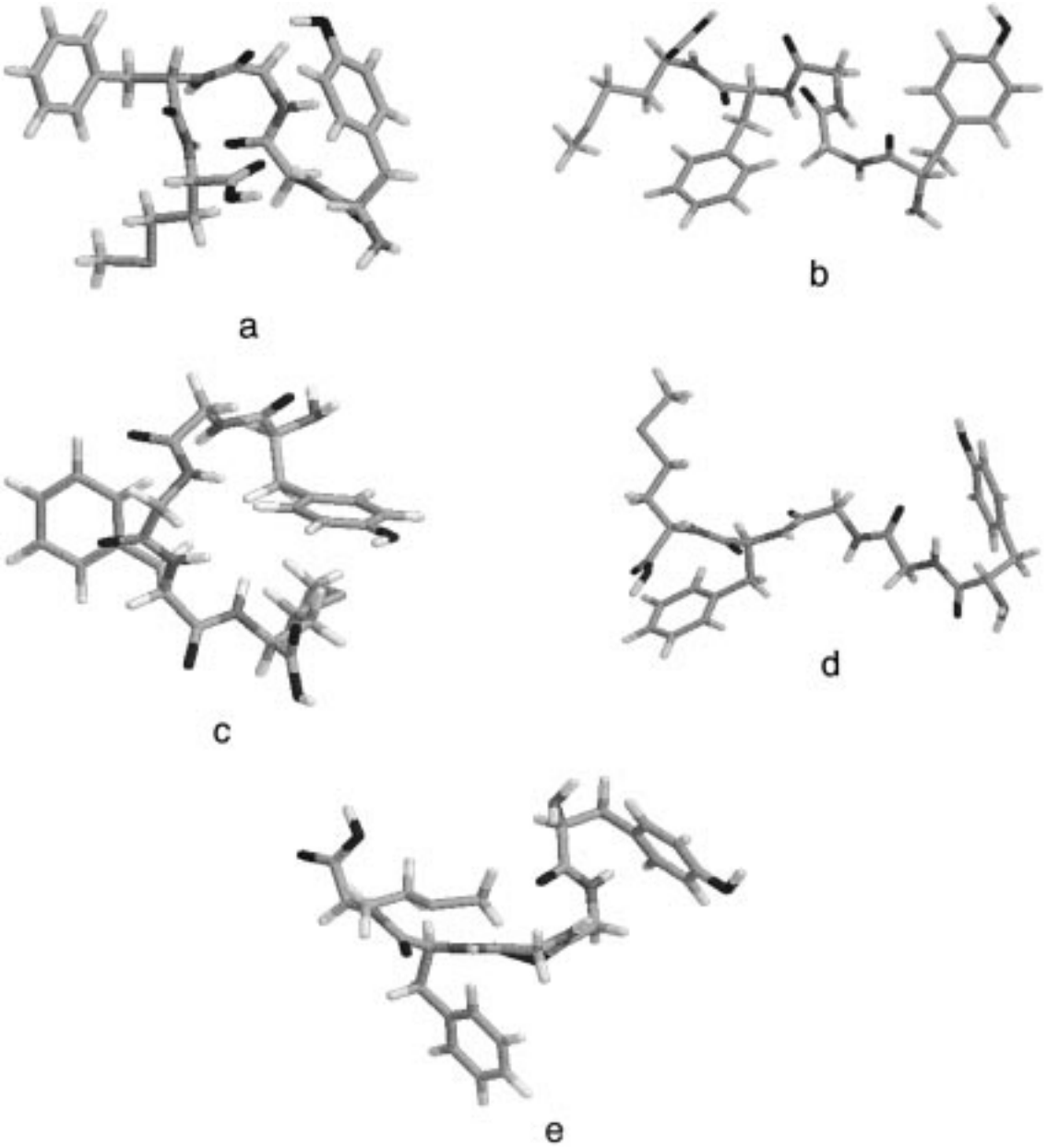

Figure 1. Initial conformations for Monte Carlo simulations for Met-enkephalin used in the present work: (a) conformation 1, (b) conformation 2, (c) conformation 3, (d) conformation 4, and (e) conformation 5. These figures were created with RasMol.

system comprising a peptide and many solvent molecules, which require an enormous amount of computational effort at each simulation step.

Our algorithm for solving the RISM equations is particularly amenable to the construction of the combined program because it is far more robust and orders of magnitude faster than the conventional one.[13,14] In fact, we tried to compare our algorithm with the conventional one in terms of the speed for Met-enkephalin. However, in the conventional one severe instability was unavoidably observed despite our every effort in setting the initial values of the iteration variables, and the convergence could never be achieved. Then, we made the comparison for acrolein with eight atomic sites and found that our algorithm is over 100 times faster than the conventional one.

In ref 14 we showed that the results of analyses on the conformational stability in water are qualitatively the same in both of the un-ionized and zwitterion cases. In the present article, we thus consider the un-ionized Met-enkephalin. The four conformations considered in our previous paper[14] and a conformation generated by assigning the torsion angles randomly (conformation 5) are used as the initial ones for the conformational sampling (these conformations are shown in Figure 1). Conformation 1 is the lowest-energy conformation in gas-phase already determined[6] and has hydrogen bonding between H^{η} of Tyr^1 side chain and the carbonyl oxygen of Gly^3 backbone. Conformation 1 is rather compact because of this hydrogen bonding. In conformation 2 the five carbonyl oxygens are not far apart. Conformation 3 is a conformation we have obtained from the backbone dihedral angles given in ref 17. These angles were determined from NMR experiments for Met-enkephalin in an aqueous solution with the presence of 50 mM sodium dodecyl sulfate[17] (SDS) (the critical micellar concentration is 8.3 mM). Conformation 4 is almost fully extended and is similar to those determined in the NMR experiments for Met-enkephalin in an aqueous solution. Conformational energies for the five conformations are −12.0, 12.2, −2.5, 0.8, and 877.2 kcal/mol, respectively. Conformation 5 has a rather large conformational energy because of a van der Waals contact. The

Table 1. Solvation Free Energies (kcal/mol) of Met-Enkephalin with Four Different Conformations in Four Different Cases[a]

conf	case 1	case 2	case 3	case 4
1	223.9	180.4	216.4	196.8
2	222.1 (−1.8)	176.8 (−3.6)	208.8 (−7.6)	178.0 (−18.8)
3	229.9 (+6.0)	187.7 (+7.3)	228.5 (+12.1)	202.7 (+5.9)
4	220.7 (−3.2)	173.2 (−7.2)	201.0 (−15.4)	176.8 (−20.0)

[a] Case 1: the solvent is the simple, repulsive-potential system (see eq 6). Case 2: the solvent is a Lennard-Jones potential system (ϵ = 0.156 kcal/mol and σ = 0.316 nm). Case 3: the solvent is the SPC/E water but all the site-charges of Met-enkephalin are set to zero. Case 4: the solvent is the SPC/E water and the full values are assigned to the site-charges.

most important matter is that the five conformations differ from one another significantly (more detailed comparisons are given in ref 14).

Results and Discussion

Solvent Effects. Before we discuss the results obtained using our combined program, we examine effects of solvent species on the solvation free energies of Met-enkephalin with conformations 1−4. Four different cases are tested to study the solvent effects. In case 1 the solvent is the simple, repulsive-potential system described above. In case 2 the solvent is a Lennard-Jones potential system. The values of ϵ and σ in case 2 are those for oxygen of the SPC/E water[16] (ϵ = 0.156 kcal/mol and σ = 0.316 nm). The solvent particles interact with atomic sites of Met-enkephalin through Lennard-Jones potential (i.e., eq 4 with $q_A = 0$ and $q_B = 0$). The solvents in cases 1 and 2 are categorized as simple solvents. In case 3 the solvent is the SPC/E water, but all the site-charges of Met-enkephalin are set to zero. In case 4 the solvent is the SPC/E water, and the full values are assigned to the site-charges. The electrostatic interaction between Met-enkephalin and the solvent is considered only in case 4. The solvation free energies calculated by the RISM theory in the four cases are summarized in Table 1, where the values in parentheses represent those relative to that of Met-enkephalin with conformation 1.

It is observed from Table 1 that in cases 3 and 4 the differences in the solvation free energy among the four conformations are considerably large. In cases 1 and 2, on the other hand, they are much smaller though the solvation free energies themselves are very large. Since the differences in case 3 are larger than those in cases 1 and 2, the great dependence (of the solvation free energy) on the peptide conformation in the water cases is not ascribed to the electrostatic interaction between the peptide and the solvent. It is related to reorganization of the hydrogen bonding among water molecules near the peptide. Conformation 1, the lowest-energy conformation in gas phase, is still the most stable in the solvents in cases 1 and 2. On the other hand, conformation 4 is the most stable in cases 3 and 4. Thus, water is clearly distinguished from the simple solvents. We consider cases 1 and 4 hereafter. The qualitative aspects of the conclusions drawn for case 1 should also be true for case 2.

Case 1. The solvation free energies, conformational energies, and total energies (kcal/mol) of Met-enkephalin with conformations 1−4 in case 1 are summarized in Table 2. The variation in the solvation free energy (the maximum difference is 9.2 kcal/mol) is much smaller than that in the conformational energy (the maximum difference is 24.2 kcal/mol). This is suggestive that in case 1 the conformational energy dominates in determining the conformational stability of the peptide.

Table 2. Solvation Free Energies ($\Delta\mu_S$), Conformational Energies (E_C), and Total Energies ($E_T = \Delta\mu_S + E_C$) of Met-Enkephalin with Four Different Conformations in Case 1

conf	$\Delta\mu_S$ (kcal/mol)	E_C (kcal/mol)	E_T (kcal/mol)
1	223.9	−12.0	211.9
2	222.1	12.2	234.3
3	229.9	−2.5	227.4
4	220.7	0.8	221.5

Table 3. Solvation Free Energies ($\Delta\mu_S$), Conformational Energies (E_C), and Total Energies ($E_T = \Delta\mu_S + E_C$) of Met-Enkephalin with the Lowest-Energy Conformation in Case 1 Found by the Combined Program[a]

run	$\Delta\mu_S$ (kcal/mol)	E_C (kcal/mol)	E_T (kcal/mol)
1-1-200	223.9	−12.0	211.9
1-1-1000	223.9	−12.0	211.9
1-1-2000	223.9	−12.0	211.9
1-1-4000	223.9	−12.0	211.9
1-2-200	219.5	2.0	221.5
1-3-200	220.1	3.6	223.7
1-4-200	219.3	0.8	220.1
1-5-100	219.6	1.5	221.1
1-5-200	222.8	−2.2	220.6
1-5-1000	222.3	−6.7	215.6
1-5-2000	218.8	−1.7	217.1

[a] "I-J-K" represents that case I is treated, the initial conformation is conformation J, and the number of total Monte Carlo sweeps is K. One sweep updates all of the 19 torsion angles in Met-enkephalin, and 19 different conformations are sampled per sweep. In the first four runs ($J = 1$), the lowest-energy conformation is the initial conformation, i.e., conformation 1.

Next, we consider the results obtained using our combined program. A total of 11 different runs have been performed. "1-5-1000", for example, represents that case 1 is treated, conformation 5 is the initial one, and the number of total MC sweeps is 1000. The number of total sweeps in each run is in the range 100−4000. One MC sweep updates all of the 19 torsion angles in Met-enkephalin (i.e., 19 different conformations are sampled and the full RISM equations are solved 19 times per MC sweep). The solvation free energies, conformational energies, and total energies (kcal/mol) of the lowest-energy conformations obtained in these runs are summarized in Table 3. A significant result is that we have never found a conformation whose total energy is lower than that of conformation 1.

The lowest-energy conformations obtained in the runs, "1-5-100", "1-5-200", "1-5-1000", and "1-5-2000", are compared in Figure 2. Those obtained in "1-1-200", "1-2-200", "1-3-200", "1-4-200", and "1-5-200" are compared in Figure 3. We note that the lowest-energy conformation obtained in "1-1-200" is conformation 1. The lowest-energy conformations in "1-5-100" and "1-5-1000" have hydrogen bonding between H^{η} of Tyr^1 side chain and the carbonyl oxygen of Gly^3 backbone as conformation 1. The conformations shown in these figures are qualitatively different. Besides, there are significant differences in the total energy among these conformations as observed in Table 3: the maximum difference is 11.8 kcal/mol. Hence, the numbers of total MC sweeps in the runs (100−4000) are still too small, although there is a trend that the search proceeds toward conformations which are similar to conformation 1. As a conclusion, the characteristics observed in case 1 are similar to those in the gas-phase case, where the probability of finding the lowest-energy conformation (conformation 1) is ∼15% in MC simulated annealing with 10 000 total MC sweeps.[8] Conformation 1 is presumably one of the most stable conformations in case 1.

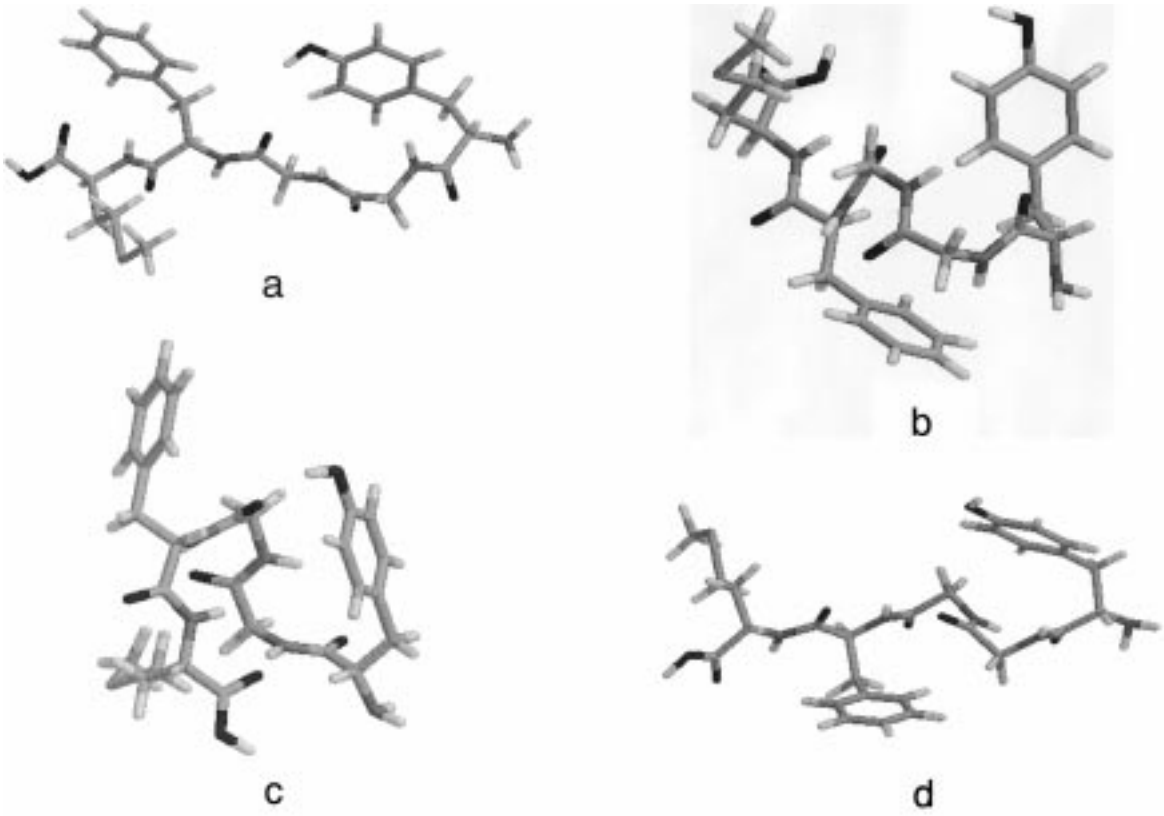

Figure 2. The lowest-energy conformations of Met-enkephalin obtained in the runs for case 1: (a) "1-5-100", (b) "1-5-200", (c) "1-5-1000", and (d) "1-5-2000". "I-J-K" represents that case I is treated, the initial conformation is conformation J, and the number of total Monte Carlo sweeps is K. In case 1 the solvent is the simple, repulsive-potential system (see eq 6). These figures were created with RasMol.

The algorithm for solving the RISM equations is a judicious hybrid of the Picard and Newton−Raphson methods.[13] One of the great advantages of the algorithm is that the Jacobian matrix for the Newton−Raphson method is treated as part of the input data. Furthermore, we have found that only one matrix calculated for a conformation (conformation 4 is chosen in the present study) can be used for all the runs of the combined program. Since the construction and LU decomposition[13] of the matrix is the most time-consuming part of the algorithm, this finding will be very useful in future studies treating larger polypeptides or proteins. The average computation time required for solving the RISM equations for one conformation is only ∼0.3 min.

Case 4. The solvation free energies, conformational energies, and total energies (kcal/mol) of Met-enkephalin with four different conformations in case 4 are summarized in Table 4. The variation in the solvation free energy (the maximum difference is 25.9 kcal/mol) is comparable with that in the conformational energy (the maximum difference is 24.2 kcal/mol). This is suggestive that in case 4 the conformational stability of the peptide is greatly influenced by the solvation free energy.

We then consider the results obtained using our combined program. A total of eight different runs have been performed. The number of total MC sweeps in each run is in the range 100−400. Despite the small numbers of conformations sampled, the lowest-energy conformations obtained in these runs look qualitatively similar. To illustrate this, the lowest-energy conformations obtained in the runs, "4-5-100", "4-5-200", and "4-5-400", are compared in Figure 4, and those in "4-1-200", "4-4-200", and "4-5-200" are in Figure 5. They are all considerably extended (they are also similar to conformation 4). Their total energies are always smaller than those for conformations 1, 2, and 3. Moreover, the total energies of the lowest-energy conformations are almost the same (176.1 ± 0.7 kcal/mol) except in the two runs starting from conformation 5 (i.e., in "4-5-100" and "4-5-400"). Conformation 5 is randomly generated, and it has a rather large conformational energy with a van der Waals contact. As long as a conformation without such bad contacts (e.g., conformations 1−4) is chosen as the initial one, low-energy conformations in case 4 can be found even in 200 total MC sweeps.

We have found that as in case 1 only one Jacobian matrix calculated for a conformation (conformation 4) can be used for the RISM part in all the runs of the combined program. The average computation time required for solving the RISM equations for one conformation is only ∼1.0 min.

Comparison of Results Obtained in Gas Phase, Case 1, and Case 4. For comparison, we have performed seven runs in gas phase where the initial conformations and numbers of total MC sweeps are identical to those in the seven runs for case 1 (the four runs where the lowest-energy conformation is conformation 1 are not considered), respectively. The seven conformations obtained as the lowest-energy ones in gas phase and conformation 1 (i.e., a total of eight conformations) are compared in Figure 6(a). They are superposed so that backbone structures are best-fit in terms of the root-mean-square distance (RMSD). The eight conformations obtained as the lowest-energy conformations in case 1 (including conformation 1) are superposed in a similar manner as shown in Figure 6b. The superposition of the eight lowest-energy conformations in case 4 is shown in Figure 6c. The maximum RMSDs calculated in gas phase, case 1, and case 4 are 0.30, 0.42, and 0.24 nm, respectively. It is observed that backbone structures in case 4 are much better converged than in gas phase and in case 1.

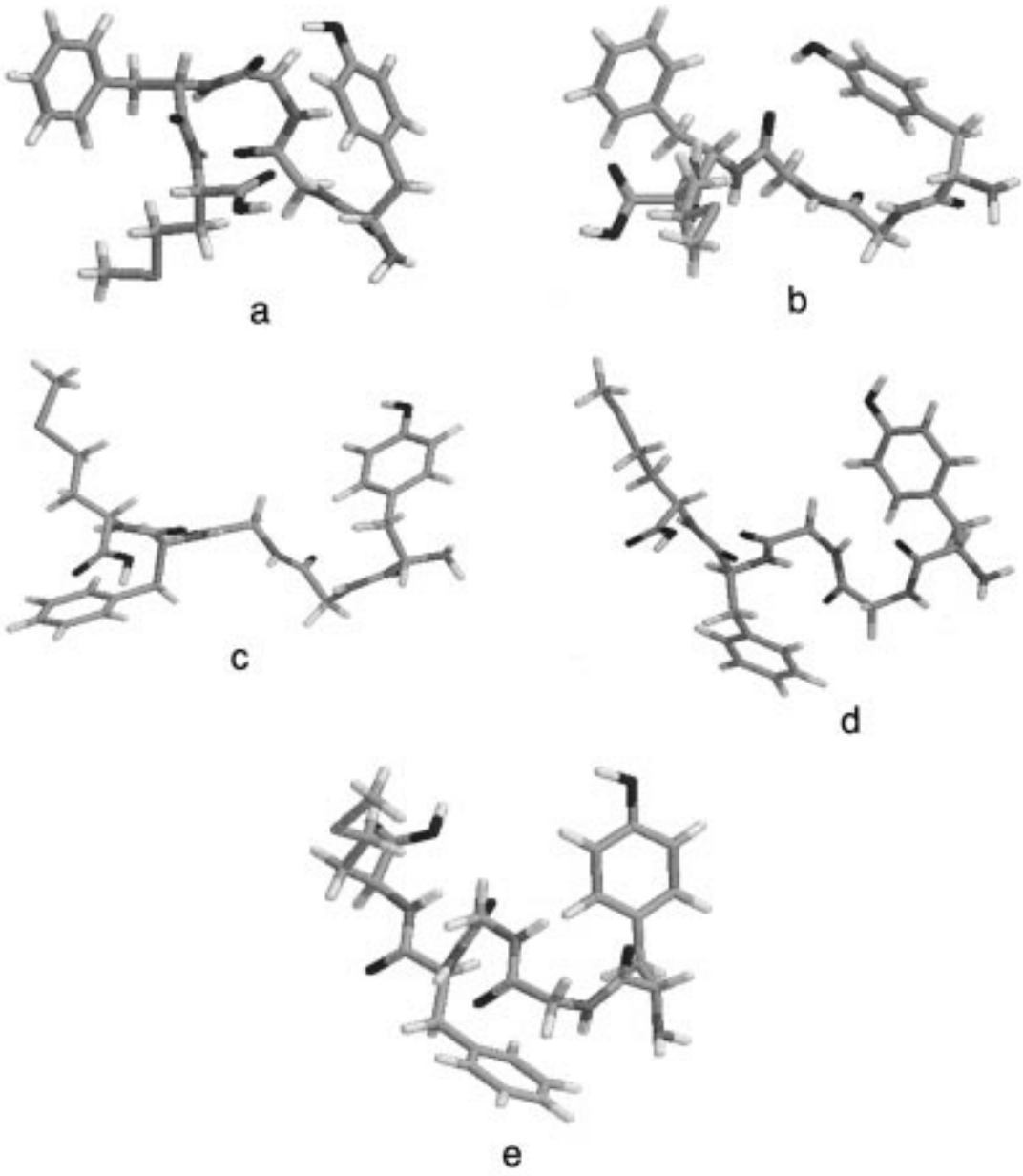

Figure 3. The lowest-energy conformations of Met-enkephalin obtained in the runs for case 1: (a) "1-1-200", (b) "1-2-200", (c) "1-3-200", (d) "1-4-200", and (e) "1-5-200". These figures were created with RasMol.

Table 4. Solvation Free Energies ($\Delta\mu_S$), Conformational Energies (E_C), and Total Energies ($E_T = \Delta\mu_S + E_C$) of Met-Enkephalin with Four Different Conformations in Case 4

conf	$\Delta\mu_S$ (kcal/mol)	E_C (kcal/mol)	E_T (kcal/mol)
1	196.8	−12.0	184.8
2	178.0	12.2	190.2
3	202.7	−2.5	200.2
4	176.8	0.8	177.6

Table 5. Solvation Free Energies ($\Delta\mu_S$), Conformational Energies (E_C), and Total Energies ($E_T = \Delta\mu_S + E_C$) of Met-Enkephalin with the Lowest-Energy Conformation in Case 4 Found by the Combined Program[a]

run	$\Delta\mu_S$ (kcal/mol)	E_C (kcal/mol)	E_T (kcal/mol)
4-1-100	159.1	16.5	175.6
4-1-200	164.5	12.3	176.8
4-2-200	165.8	9.6	175.4
4-3-200	164.9	11.4	176.3
4-4-200	164.0	12.5	176.5
4-5-100	172.4	7.7	180.1
4-5-200	165.2	10.6	175.8
4-5-400	163.1	17.2	180.3

[a] "I-J-K" represents that case I is treated, the initial conformation is conformation J, and the number of total Monte Carlo sweeps is K.

Comparison of Results Obtained in Case 4 and NMR Experiments. The conformations of Met-enkephalin in an aqueous solution (buffered to pH 3.87 using CH_3 COONa at an ionic concentration of 0.05 M) were already determined from the NMR experiments.[17] It is true that effects due to the presence of CH_3 COONa is unknown and need to be investigated in further studies. However, we considered a dipeptide in a 1 M NaCl solution and concluded that the conformational stability of very small peptides is not affected by the salt addition[21] (the solvation free energy increases, but the degree of the increase is not appreciably dependent on the conformation). Hence, it is worthwhile to compare the NMR results with those in the present study.

(21) Kinoshita, M.; Okamoto, Y.; Hirata, F. *J. Comput. Chem.*, submitted.

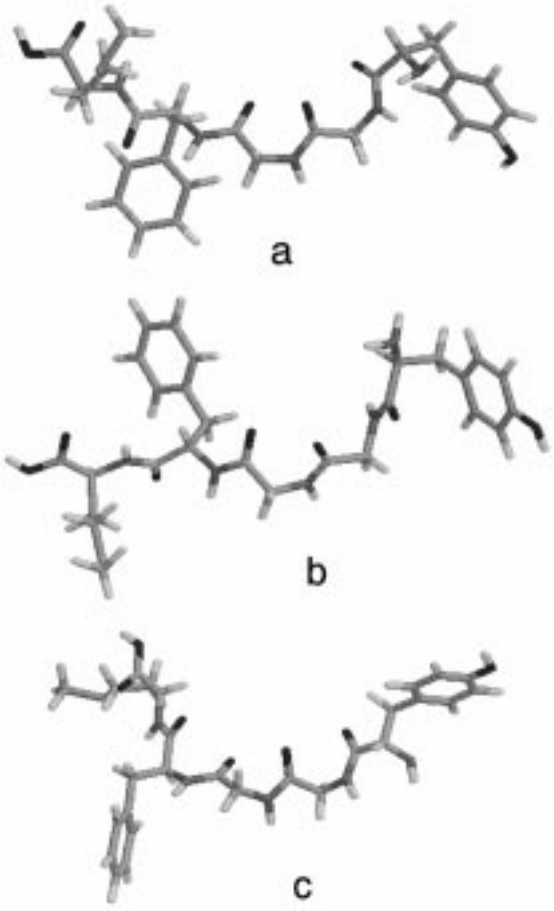

Figure 4. The lowest-energy conformations of Met-enkephalin obtained in the runs for case 4: (a) "4-5-100", (b) "4-5-200", and (c) "4-5-400". In case 4 the solvent is the SPC/E water and the full values are assigned to the site-charges of Met-enkephalin. These figures were created with RasMol.

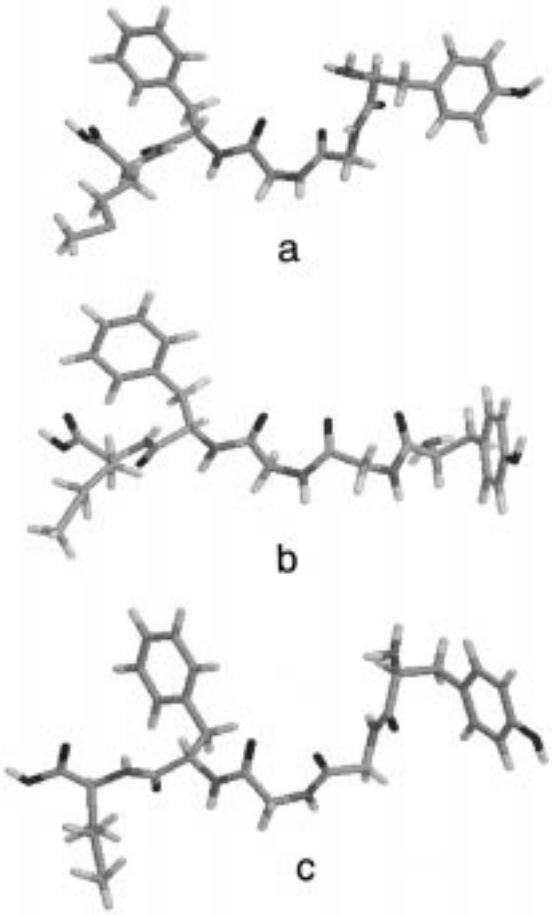

Figure 5. The lowest-energy conformations of Met-enkephalin obtained in the runs for case 4: (a) "4-1-200", (b) "4-4-200", and (c) "4-5-200". These figures were created with RasMol.

In ref 17, only five of the 20 conformations obtained in the experiments are selected, and only eight of the 19 torsion angles are given for the five conformations. Moreover, most of the eight torsion angles vary remarkably from conformation to conformation. Therefore, it is meaningless to make a quantitative comparison in terms of the torsion angles. However, it is definite that the conformations obtained in the NMR experiments are characterized by almost fully extended backbone structure with large fluctuations in side-chain structure. The eight conformations shown in Figure 6c are quite similar to the experimentally determined ones (see Figure 2 of ref 17) when they are visually inspected. Moreover, the RMSDs for the conformations in Figure 2 of ref 17 are in the range 0.12−0.22 nm, and those in Figure 6c are 0.11−0.24 nm: the conformations obtained in case 4 are as well converged as those determined in the NMR experiments. The set of stable conformations in water can be found in much less total MC sweeps than in gas phase and in simple solvents, and they are in qualitatively good agreement with the experimental observations. The intramolecular hydrogen bonding seen in the lowest-energy conformation in gas phase is no more present in the stable conformations in water. This is because H^{η} of Tyr^1 side chain and the carbonyl oxygen of Gly^3 backbone form hydrogen bonding with water−oxygen and water−hydrogen, respectively. As a consequence, the conformations in water are almost fully extended and in qualitative accord with the NMR results. With the intramolecular hydrogen bonding, the backbone structure would be less extended.

Possible Relevance to Levinthal's Paradox. In summary, we have found that simulations in water converge much faster than those in gas phase and in simple solvents. This is because variations in the solvation free energy among different peptide conformations in water are considerably large. Most of the conformations with larger solvation free energies are strongly rejected by water, and the number of probable (low-energy) conformations is drastically reduced, which is suggestive that a peptide is forced to take conformations favored by water quite rapidly. Levinthal[22] pointed out that the time for a random search of all possible conformations would be unrealistically long even for a small protein. "Water" might lead us to a solution to this paradox. Of course, further studies are needed to prove the validity of our suggestion for larger peptides or proteins, and work in this direction is in progress.

Conclusion

We have developed a computer program combining Monte Carlo simulated annealing and the full RISM theory. Our robust algorithm for solving the RISM equations, which is orders of magnitude faster than the conventional one, is employed (we emphasize that no approximate treatment is used to accelerate the RISM calculation). After sampling many conformations of peptides in accordance with the simulated annealing technique,

(22) (a) Levinthal, C. *J. Chim. Phys.* **1968**, *65*, 44−45. (b) Wetlaufer, D. B. *Proc. Natl. Acad. Sci. U.S.A.* **1973**, *70*, 697−701.

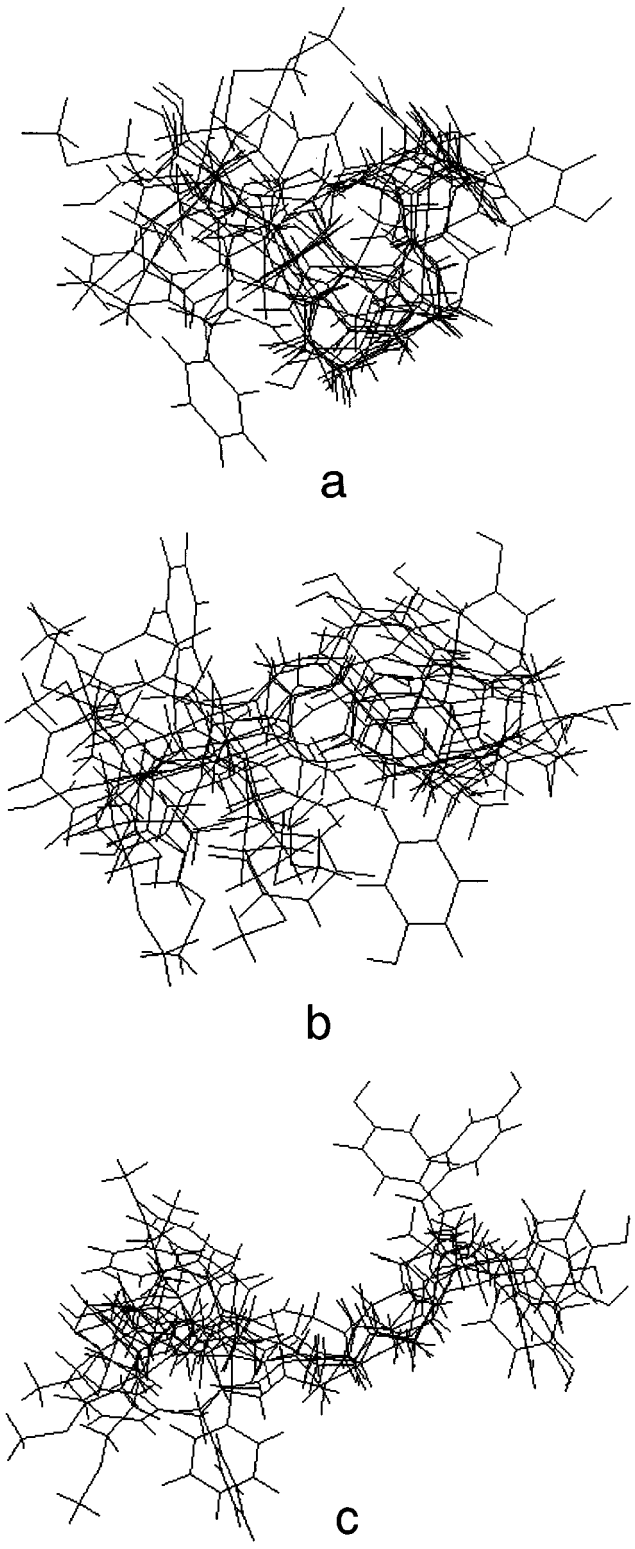

Figure 6. Superposition of the eight conformations of Met-enkephalin obtained as the lowest-energy ones in (a) gas phase, (b) case 1, and (c) case 4. These figures were created with RasMol. Figure 6c should be compared with Figure 2 in ref 17.

the combined program finds the conformation with the minimum total energy. The total energy is the sum of the conformational energy and the solvation free energy, and the latter is calculated by the RISM theory. The effectiveness of the program has been demonstrated for Met-enkephalin in water and in a simple, repulsive-potential system. We have found that in the RISM part only one Jacobian matrix calculated for a conformation can be used throughout the conformational sampling in all the runs tested, leading to considerable saving of computation time. The RISM equations must be solved a number of times, but this can be performed with moderate computational effort on a workstation. The algorithm for solving the RISM equations has never failed to give convergence.

Since in simple solvents the solvation free energy remains roughly constant against conformational changes, the conformational energy is more important. The most stable conformation in gas phase is still considerably stable in the simple solvents. In water, on the other hand, variations in the solvation free energy among different conformations are considerably large. Most of the conformations are strongly rejected by water, and the number of probable conformations is drastically reduced, which is suggestive that Met-enkephalin is forced to take conformations favored by water quite rapidly. This result is important in the following two respects: (i) water plays essential roles in stabilizing particular conformations of peptides, and (ii) the number of conformations to be sampled in simulations is drastically reduced in water, which could remove computational bottlenecks expected for much larger polypeptides or proteins. It has been shown that in water there are a set of different conformations of Met-enkephalin having almost the same total energies, nearly the lowest values. These conformations exhibit characteristics of almost fully extended backbone structure with large fluctuations in side-chain structure, which are in qualitatively good agreement with the observations in the recent NMR experiments (these conformations are quite different from the most stable conformation in gas phase).

In water, the solvation free energy is considerably dependent on the peptide conformation even when all the site-charges of the peptide are set to zero. Hence, the water structure itself should be responsible for the large dependency. Compared with simple (i.e., nonpolar) solvents, the structure of water is orientational due to the hydrogen-bonding network, and this presumably causes the large dependency. In the scaled particle theory[23] for elucidating the solvent effects, water molecules are simply treated as sufficiently small hard spheres which are similar to the simple, repulsive-potential particles considered in the present study. This theory assumes that the characteristic differences between a nonpolar solvent and water are not due to the water structure but due to the comparatively small size of water molecules[24] and has been successful for nearly spherical solutes.[23] However, our result is suggestive that such an assumption is not applicable to complicated molecules such as peptides which can take a variety of conformations. The purpose of the combined program is to find a set of stable conformations with sufficiently low total energy. It is not necessary to find the global minimum. In the simple, repulsive-potential system it is obvious that such a set of conformations has not been found yet (the maximum difference in the total energies of the eight conformations obtained is 11.8 kcal/mol). In water, however, it can be concluded that such a set of conformations has been found for the following reasons: the maximum difference in the total energy is only 1.4 kcal/mol; the root-mean-square distance for the eight conformations is 0.11−0.24 nm and that for the NMR conformations is 0.12−0.22 nm, and hence the conformations obtained in water are as

(23) (a) Pierotti, R. A. *J. Phys. Chem.* **1965**, *69*, 281−288. (b) Pierotti, R. A. *Chem. Rev.* **1976**, *76*, 717−726. (c) Shoor, S. K.; Gubbins, K. E. *J. Phys. Chem.* **1969**, *73*, 498−505. (d) Masterton, W. L.; Lee, T. P. *J. Phys. Chem.* **1970**, *74*, 1776−1782.

(24) Pohorille A.; Pratt, L. R. *J. Am. Chem. Soc.* **1990**, *112*, 5066−5074.

well converged as those from the NMR experiments. In the combined program, the effects due to peptide dynamics are not taken into account, i.e., the conformational entropy is neglected. In the present study for Met-enkephalin, however, the stable conformations obtained are almost fully extended. Therefore, incorporation of the conformational entropy will even enhance the stability of these conformations. For larger polypeptides and proteins, the multicanonical approach[2] should be employed instead of simulated annealing, to account for the effects of dynamics. The approach allows us to obtain the ensemble-averaged conformation rather than the lowest-energy conformation. Work in this direction is in progress.

Acknowledgment. The present study was supported by the Joint Studies Program (1996) of the Institute for Molecular Science, "Drastic Reduction in Computer Time of Calculating Solvation Free Energy for Protein Using RISM Theory" and also by Grants-in-Aid for Scientific Research from the Japanese Ministry of Education, Science, Sports and Culture.

JA972048R