THEORETICAL STUDIES OF CONFORMATIONS OF OLIGOPEPTIDES BY EFFICIENT SIMULATION METHODS

A Thesis
Presented to the Department of Functional Molecular Science
School of Mathematical and Physical Science
The Graduate University for Advanced Studies
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Science

by

Ayori Mitsutake

September 1999

To my parents, Naoto and Miyoko Mitsutake

Acknowledgments

My most heartfelt thanks go to my thesis advisor, Professor Yuko Okamoto, for his patient guidance and constant encouragement. I would like to thank Professors Ulrich H.E. Hansmann, Fumio Hirata, Masayuki Irisa, and Masahiro Kinoshita for valuable and helpful discussions and suggestions. I am greatly indebted to Drs. Tsutomu Ikegami, Masato Masuya, Takeshi Nishikawa, and Yuji Sugita for useful discussions and their help in maintenance of workstations. I wish to express my gratitude to all the members of the Okamoto and Hirata Groups for stimulating discussions and encouragement. I am grateful to the members of the IMS theory groups for their generous support. I wish to thank the faculty and staff members of the Graduate University for Advanced Studies for their kindness. I am most thankful to my parents Naoto and Miyoko Mitsutake, my brother Kazumasa Mitsutake, and my uncle Takaichiro Tomonaga for their constant encouragement and support. Finally, I am grateful to the Japan Society for the Promotion of Science for financial aid.

Contents

Chapter 1

General Introduction

Proteins fold into the unique three-dimensional structures (tertiary structures) from random-coil states in the time scale of miliseconds to minutes. Although the structures of proteins seem to depend on the environment in the cell where they are synthesized, it was shown by experiments *in vitro* that the unfolded protein can refold back into their native conformations once the denaturants are removed [1]. This implies that the tertiary structures of proteins are determined by their amino-acid sequence information and that the structure of a native protein corresponds to the conformation with the lowest free energy. Much efforts have been invested to find such conformations for peptide and protein molecules without complete success (for a review, see, for instance, Ref. [2]). There are two difficulties in predicting the tertiary structure of a protein. One is how to give the accurate energy function of the system. Especially, the effects of solvent surrounding a protein are non-trivial to incorporate, because the number of solvent molecules that have to be considered is very large. The other difficulty is due to the fact that there exist a huge number of local minima in the potential energy function, which makes it prohibitively difficult for conventional simulation methods to search for the global minimum.

There are many ways to incorporate solvent effects. One of the simplest and least time-consuming methods is probably to use the distance-dependent dielectric function [3]. Simulations with this method were performed in Refs. [4, 5]. Note that it only involves the electrostatic interactions. Other solvent contributions (namely, hydrophobic interactions and Lennard-Jones interactions between protein and solvent) are not included. Another commonly used term that represents solvent contributions is the term proportional to the solvent accessible surface area of protein molecule [6, 7]. It is thus implicitly

assumed that the range of protein-solvent interactions is confined to the first solvent hydration shell. The validity of these assumptions were supported by the experiments which showed that the transfer free energies of many nonpolar molecules vary linearly with the solvent-accessible surface area [8, 9]. The proportionality constants are determined from experimental data of small molecules, and hence this solvation term, in principle, includes all the contributions from solvent. In this thesis we used the parameters of Ref. [10].

There are a few methods for including solvent effects that are based on the statistical mechanics of liquids. One of such approaches is RISM [11, 12]. Recently, this algorithm was used in simulations of protein folding [13]-[15]. This method is very powerful, but computationally demanding. The scaled particle theory [16, 17] is another method based on the statistical mechanics of liquids, which is known to give a good account of the hydration free energy of a non-polar solute. The method requires much less computation time compared to the RISM method; thus, larger molecules can be treated. This method can calculate cavity formation free energy term. Then together with the electrostatic interactions obtained by solving Poisson-Boltzmann equations [18] and Lennard-Jones interactions between protein and solvent, this theory can give an accurate solvation free energy. Another accurate but time-consuming method is to include explicit solvent molecules directly in simulations.

As for the simulation algorithms, we adopt Monte Carlo simulated annealing [19] and multicanonical algorithm [20]. Simulated annealing [19] is based on the process of crystal making. Namely, by starting a simulation at a sufficiently high temperature (much above the melting temperature), one lowers the temperature gradually during the simulation until it reaches the global-minimum-energy state (crystal). The application of Monte Carlo simulated annealing to the prediction of protein tertiary structures was proposed almost a decade ago by a few groups [21]-[25]. Since then this method has been extensively used in the protein folding and structure refinement problems (for reviews, see Refs. [26, 27]).

While a regular Monte Carlo method generates states according to the canonical distribution, multicanonical algorithm [20] generates states so that a one-dimensional random walk in energy space is realized. The application of multicanonical algorithm to the pro-

tein folding problem was proposed several years ago [28]. Since then there have been many works based on this method and its variants in protein and related systems [29]-[47] (for reviews, see Refs. [27, 48]).

In this thesis, we discuss the folding problem of oligopeptides with all-atom models. The organization of the thesis is as follows. In Chapter 2, we explain the energy function and the sampling algorithms that we used. In Chapter 3, we present the results of Monte Carlo simulated annealing simulations of a penta peptide, Met-enkephalin, in gas phase and in a model solvent. We estimate a rigorous cavity-formation term in the solvation free energy by the extended scaled particle theory. This is the first attempt to combine Monte Carlo simulated annealing and the extended scaled particle theory. In Chapter 4, we give the results of a multicanonical simulation of Met-enkephalin in gas phase and demonstrate the effectiveness of the multicanonical simulation algorithm. In Chapter 5, we study helix-coil transitions of amino-acid homo-oligomers of length 10 (homo-alanine, homo-valine, and homo-glycine) in aqueous solution by multicanonical Monte Carlo simulations. The solvent effects were included by the term that is proportional to the solvent-accessible surface area. Finally, Chapter 6 is devoted to conclusions.

Bibliography

[1] C. J. Epstain, R. F. Goldberger, and C. B. Anfinsen, *Cold Spring Harbor Symp. Quant. Biol.* **28**, 439 (1963).

[2] M. Vásquez, G. Némethy, and H. A. Scheraga, *Chem. Rev.* **94**, 2183 (1994).

[3] a) B. E. Hingerty, R. H. Ritchie, T. Ferrell, and J. E. Turner, *Biopolymers* **24**, 427 (1985).

b) J. Ramstein and R. Lavery, *Proc. Natl. Acad. Sci. USA* **85**, 7231 (1988).

[4] V. Daggett, P. A. Kollman, and I. D. Kuntz, *Biopolymers* **31**, 285 (1991).

[5] Y. Okamoto, *Biopolymers* **34**, 529 (1994).

[6] C. Chothia, *Nature* **248**, 338 (1974).

[7] D. Eisenberg and A. D. McLachlan, *Nature* **319**, 199 (1986).

[8] R. B. Hermann, *J. Phys. Chem.* **76**, 2754 (1972).

[9] J. A. Reynolds, D. B. Gilbert, and C. Tanford, *Proc. Natl. Acad. Sci. USA* **71**, 2925 (1974).

[10] T. Ooi, M. Oobatake, G. Némethy, H. A. Scheraga, *Proc. Natl. Acad. Sci. USA* **84**, 3086 (1987).

[11] D. Chandler and H. C. Andersen, *J. Chem. Phys.* **57**, 1930 (1972).

[12] F. Hirata and P. J. Rossky, *Chem. Phys. Lett.* **83**, 329 (1981).

[13] M. Kinoshita, Y. Okamoto, and F. Hirata, *J. Comput. Chem.* **18**, 1320 (1997).

[14] M. Kinoshita, Y. Okamoto, and F. Hirata, *J. Chem. Phys.* **107**, 1586 (1997).

[15] M. Kinoshita, Y. Okamoto, and F. Hirata, *J. Am. Chem. Soc.* **120**, 1855 (1998).

[16] H. Reiss, H. L. Fish, and J. L. Lebowitz, *J. Chem. Phys.* **31**, 369 (1959).

[17] R. A. Pierotti, *Chem. Rev.* **76**, 717 (1965).

[18] a) B. Honig and A. Nicholls, *Science* **268**, 1144 (1995).
b) H. Nakamura, *Quart. Rev. of Biophys.* **29**, 1 (1996).

[19] S. Kirkpatrick, C. D. Gelatt, Jr., and M. P. Vecchi, *Science* **220**, 671 (1983).

[20] a) B. A. Berg and T. Neuhaus, *Phys. Lett.* **B267**, 249 (1991).
b) B. A. Berg and T. Neuhaus, *Phys. Rev. Lett.* **68**, 9 (1992).

[21] S. R. Wilson, W. Cui, J. W. Moskowitz, and K. E. Schmidt, *Tetrahedron Lett.* **29**, 4373 (1988).

[22] A. T. Brünger, *J. Mol. Biol.* **203**, 803 (1988).

[23] M. Nilges, G. M. Clore, and A. M. Gronenborn, *FEBS Lett.* **229**, 317 (1988).

[24] H. Kawai, T. Kikuchi, and Y. Okamoto, *Protein Eng.* **3**, 85 (1989).

[25] C. Wilson and S. Doniach, *PROTEINS: Struct. Funct. Genet.* **6**,193 (1989).

[26] S. R. Wilson and W. Cui, In: *The Protein Folding Problem and Tertiary Structure Prediction,* K. M. Merz, Jr. and S. M. Le Grand, eds, (Birkhäuser, Boston, 1994) p. 43.

[27] Y. Okamoto, *Recent Res, Devel. in Pure & Applied Chem.* **2**, 1 (1998).

[28] U. H. E. Hansmann and Y. Okamoto, *J. Comp. Chem.* **14**, 1333 (1993).

[29] U. H. E. Hansmann and Y. Okamoto, *Physica* **A212**, 415 (1994).

[30] M. H. Hao and H. A. Scheraga, *J. Phys. Chem.* **98**, 4940 (1994).

[31] a) Y. Okamoto, U. H. E. Hansmann, and T. Nakazawa, *Chem. Lett.* **1995**, 391.
b) Y. Okamoto and U. H. E. Hansmann, *J. Phys. Chem.* **99**, 11276 (1995).

[32] A. Kidera, *Proc. Natl. Acad. Sci. USA* **92**, 9886 (1995).

[33] N. Urakami and M. Takasu, *J. Phys. Soc. Jpn.* **65**, 2694 (1996).

[34] S. Kumar, P. Payne, and M. Vásquez, *J. Comp. Chem.* **17**, 1269 (1996).

[35] U. H. E. Hansmann, Y. Okamoto, and F. Eisenmenger, *Chem. Phys. Lett.* **259**, 321 (1996).

[36] N. Nakajima, H. Nakamura, and A. Kidera, *J. Phys. Chem.* **101**, 817 (1997).

[37] H. Noguchi and K. Yoshikawa, *Chem. Phys. Lett.* **278**, 184 (1997).

[38] U. H. E. Hansmann and F. Eisenmenger, *J. Phys. Chem. B* **101**, 3304 (1997).

[39] J. Higo, N. Nakajima, H. Shirai, A. Kidera, and H. Nakamura, *J. Comp. Chem.* **18**, 2086 (1997).

[40] C. Bartels and M. Karplus, *J. Phys. Chem. B* **102**, 865 (1998).

[41] Y. Iba, G. Chikenji, and M. Kikuchi, *J. Phys. Soc. Jpn.* **67**, 3327 (1998).

[42] N. Nakajima, *Chem. Phys. Lett.* **288**, 319 (1998).

[43] M. H. Hao and H. A. Scheraga, *J. Mol. Biol.* **277**, 973 (1998).

[44] H. Shirai, N. Nakajima, J. Higo, A. Kidera, and H. Nakamura, *J. Mol. Biol.* **278**, 481 (1998).

[45] M. Schaefer, C. Bartels, and M. Karplus, *J. Mol. Biol.* **284**, 835 (1998).

[46] C. Bartels, R. H. Stote, and M. Karplus, *J. Mol. Biol.* **284**, 1641 (1998).

[47] U. H. E. Hansmann and Y. Okamoto, *J. Phys. Chem. B* **103**, 1595 (1999).

[48] U. H. E. Hansmann and Y. Okamoto, *Curr. Opin. Struct. Biol.* **9**, 177 (1999).

Chapter 2

Simulation Methods

The simulation methods that were employed in this thesis are overviewed in this chapter. In section 2.1 we summarize the energy functions of protein systems that we used in our simulations. In section 2.2 we review efficient sampling algorithms for protein folding simulations.

2.1 Potential Energy Function

The potential energy function that we used in this thesis is given by the sum of two terms: the conformational energy (E_P) of the protein molecule itself and the solvation free energy ($\mathcal{E}_{SOL}$) for the interaction of the protein with the surrounding solvent:

$$E_{TOT} = E_P + \mathcal{E}_{SOL} \ . \tag{2.1}$$

Here, the solvation effects are included not by the solvation energy of fixed coordinates of water molecules, but by the solvation free energy, because we want to estimate the average solvation effects for many configurations of water molecules around each fixed solute conformation. In the following we discuss in detail each component of the above terms that we actually used in our simulations.

2.1.1 Conformational Energy Function

The conformational energy function E_P (in kcal/mol) is given by the sum of the electrostatic term E_C, 12-6 Lennard-Jones term E_{LJ}, and hydrogen-bond term E_{HB} for all pairs of atoms in the protein molecule together with the torsion term E_{TOR} for all torsion angles:

$$\begin{aligned}
E_P &= E_C + E_{LJ} + E_{HB} + E_{TOR} \ , \\
E_C &= \sum_{(i,j)} \frac{332\ q_i q_j}{\epsilon\ r_{ij}} \ , \\
E_{LJ} &= \sum_{(i,j)} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \right) \ , \\
E_{HB} &= \sum_{(i,j)} \left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \right) \ , \\
E_{TOR} &= \sum_i U_i \left(1 + \cos(n_i \chi^i + \theta_i) \right) \ .
\end{aligned} \tag{2.2}$$

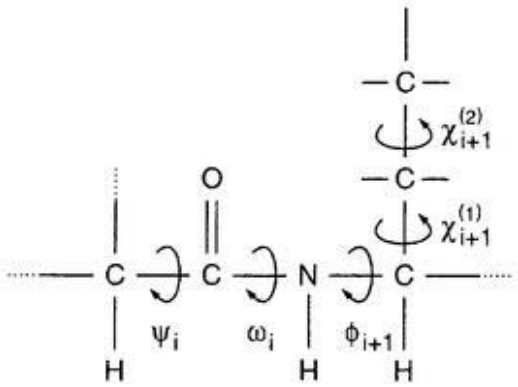


Figure 2.1: Dihedral angles in a protein molecule (ϕ_i and ψ_i in the main chain of each residue i and χ_i^j in its side chains).

Here, r_{ij} is the distance (in Å) between atoms i and j, ϵ is the dielectric constant, and χ^i is the torsion angle for the chemical bond i. Each atom is expressed by a point at its center of mass, and the partial charge q_i (in units of electronic charges) is assumed to be concentrated at that point. The factor 332 in E_C is a constant to express energy in units of kcal/mol. The parameters in the energy function as well as the molecular geometry are based on ECEPP/2 [1]. The computer code KONF90 [2, 3] is modified to incorporate various solvation effects and simulation algorithms. The dielectric constant ϵ is set equal to 2. The peptide-bond dihedral angles ω are fixed at the value 180° for simplicity. The remaining dihedral angles (ϕ_i and ψ_i in the main chain of each residue i and χ_i^j in its side chains) constitute the degree of freedom in the simulations (see Fig. 2.1). One Monte Carlo (MC) sweep consists of updating all these angles once with a Metropolis evaluation [4] for each update.

2.1.2 Solvation Free Energy

There are many ways to incorporate solvent effects. In this thesis we use the two methods described below.

Solvation Free Energy Function Based on Scaled Particle Theory

The solvation free energy of interactions between a solute molecule and solvent molecules can be divided into three contributions in general:

$$\mathcal{E}_{SOL} = \mathcal{E}_{CAV} + \mathcal{E}_{LJ} + \mathcal{E}_{EL} \ . \tag{2.3}$$

The first term is the cavity-formation term, which corresponds to the work required to create a cavity having the shape of the solute molecule in solution. The second term represents the Lennard-Jones interaction term between the solute and solvent molecules. The third term is the electrostatic interaction term (including the hydrogen-bond energy) between the solute and the solvent molecules. In this thesis, we only use the cavity formation term $\mathcal{E}_{CAV}$, which corresponds to the hydrophobic effects, because the evaluation of the other two terms are much more time-consuming.

The extended scaled particle theory [5, 6] is used to calculate the cavity-formation term ($\mathcal{E}_{CAV}$) by scaling up a solute molecule in the solvent:

$$\mathcal{E}_{CAV} = W(\lambda = 1) \ , \tag{2.4}$$

where λ is the scaling parameter, which varies from 0 (material point) to 1 (real size), and $W(\lambda)$ is the work required to dissolve the scaled solute particle to the solvent. The theory assumes that $W(\lambda)$ can be represented for all positive values of λ by

$$W(\lambda) = A + B\lambda + \frac{1}{2}C\lambda^2 + PV_c(\lambda), \quad \lambda \geq 0, \tag{2.5}$$

where $V_c(\lambda)$ is the excluded volume of the scaled solute and P is the macroscopic pressure of the solvent. The term PV_c is very small for $\lambda = 1$ under a pressure of one atmosphere, and are neglected. The coefficients A, B, and C are determined using the continuity conditions up to second derivatives at $\lambda = 0$. The explicit expressions for these coefficients are as follows [6]:

$$\begin{aligned} A &= -k_B T \ln(1 - \rho V_c(0)), \\ B &= k_B T \frac{1}{1 - \rho V_c(0)} \rho \left(\frac{\partial V_c}{\partial \lambda} \right)_{\lambda=0}, \\ C &= k_B T \frac{1}{1 - \rho V_c(0)} \rho \left(\frac{\partial^2 V_c}{\partial^2 \lambda} \right)_{\lambda=0} + k_B T \frac{1}{(1 - \rho V_c(0))^2} \rho^2 \left(\left(\frac{\partial V_c}{\partial \lambda} \right)_{\lambda=0} \right)^2, \end{aligned} \tag{2.6}$$

where k_B is the Boltzmann constant, T is the absolute temperature, and ρ is the number density of the solvent molecules. The analytical evaluations of the excluded volume ($V_c(\lambda)$) and its derivatives are carried out using the code in Ref. [7]. Together with the electrostatic term, $\mathcal{E}_{EL}$ in Eq. (2.3), obtained by solving the Poisson-Boltzmann equations [8, 9] and the Lennard-Jones term, $\mathcal{E}_{LJ}$, in Eq. (2.3), between the solute and the solvent, this theory can give a good account for the solvation free energy.

Solvation Free Energy Function Based on the Term Proportional to the Solvent-Accessible Surface Area

Although the method based on the extended scaled particle theory described in the previous subsubsection gives an accurate estimate for the cavity-formation free energy (one of the three terms in Eq. (2.3)), it is computationally very demanding if we include all the contributions to the solvation free energy (in Eq. (2.3)). Therefore, for a larger system we want to use an empirical method that gives an approximation to the whole solvation free energy. We adopt the term proportional to the solvent-accessible surface area.

In this approach the total free energy of hydration of a solute molecule is thus given by

$$\Delta E_{sol} = \sum_i \sigma_i A_i \tag{2.7}$$

where the summation extends over all groups i of atoms, and A_i is the corresponding solvent accessible surface area. The constant of proportionality σ_i represents the contribution to the solvent free energy of group i per unit accessible surface area. There exist several parameter sets in use. The validity of five of them was recently investigated and it was concluded that those of Refs. [10, 11] are valid ones [12]. In this thesis we use the parameters of Ref. [10]. Accessible surface areas are computed for seven classes of atoms or groups, listed in Table 2.1. The groups containing hydrogen atoms in Table 2.1 are treated as "united atoms". The accessible surface area A_i is obtained by the surface area of fused spheres centered at each united atom. The radius of the sphere is $R_i + R_w$, where R_w is the effective radius of the solvent molecule. Here we set R_i to van der Walls radius in Table. 2.1 and R_w to 1.4 Å[10]. For the calculation of solvent-accessible surface area,

Table 2.1: Van der Waals radii and the computed coefficients for the thermodynamic parameters.

i	Class of chemical group	van der Waals radius R_i Å	σ (kcal/mol · Å^2)
1	Aliphatic $-CH_3$, $-CH_2-$, $>CH-$	2.00	0.008
2	Aromatic $=CH-$	1.75	−0.008
3	Hydroxyl $-OH$	1.40	−0.172
4	Amide and amine $-NH-$,$-NH_2$	1.55	−0.132
5	Carboxyl and carbonyl $>C=$	1.55	0.427
6	Carboxyl and carbonyl $O=$	1.40	−0.038
7	Sulfur $-S-$ and thiol $-SH$	2.00	−0.021

we use the computer code NSOL [13], which is based on the code NSC [14]

2.2 Efficient Sampling Algorithms for Protein Folding Simulations

Once the appropriate energy function is given, we next have to employ a simulation method that does not get trapped in states of energy local minima. In this thesis we employ the Monte Carlo simulated annealing method [15] and multicanonical algorithm [16].

2.2.1 Simulated Annealing Method

In the regular canonical ensemble at a given inverse temperature $\beta \equiv 1/k_B T$, the probability weight of each state with energy E is given by the Boltzmann factor:

$$w_B(E;T) = \exp(-\beta E). \tag{2.8}$$

The probability distribution of energy is then given by

$$P_B(E;T) \propto n(E) w_B(E;T), \tag{2.9}$$

where $n(E)$ is the density of states. Since $n(E)$ is a rapidly increasing function and $w_B(E;T)$ decreases exponentially, $P_B(E;T)$ generally has a bell-like shape (see Fig. 2.2).

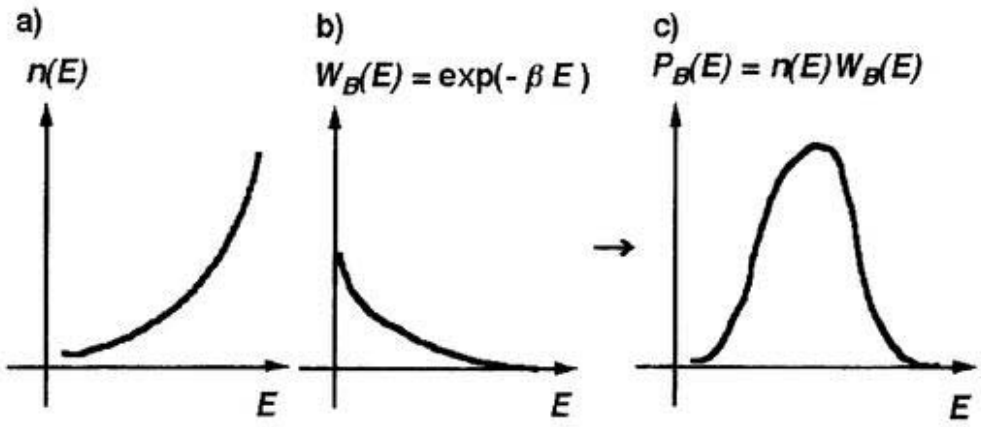


Figure 2.2: The density of states $n(E)$ (a), the Boltsmann weight factor $w_B(E;T)$ (b), and the probability distribution $P_B(E;T)$ (c) as a function of energy E.

When the temperature is high, β is small, and $w_B(E;T)$ decreases slowly with E. The probability distribution $P_B(E;T)$ then has a wide bell-shape. On the other hand, at low temperature β is large, and $w_B(E;T)$ decreases rapidly with E. The probability distribution $P_B(E;T)$ has a narrow bell-shape (and in the limit $T \to 0$ K, $P_B(T,E) \propto \delta(E-E_{GS})$, where E_{GS} is the global-minimum energy). The canonical probability distributions at different temperatures are illustrated in Fig. 2.3. However, it is very difficult to obtain canonical distributions at low temperatures with conventional simulation methods. This is because the thermal fluctuations at low temperatures are small and the simulation will certainly get trapped in states of energy local minima.

Simulated annealing [15] is based on the process of crystal making process. Namely, by starting a simulation at a sufficiently high temperature (much above the melting temperature), one lowers the temperature gradually during the simulation until it reaches the global- minimum-energy state (crystal). If the rate of temperature decrease is sufficiently slow so that thermal equilibrium may be maintained throughout the simulation, only the state with the global energy minimum is obtained (when the final temperature is 0 K).

In this thesis, we adopt the following annealing schedule: The temperature is lowered exponentially from $T_I = 1000$ K to $T_F = 150$ K. The temperature for the n-th MC sweep is given by

$$T_n = T_I \gamma^{n-1}, \tag{2.10}$$

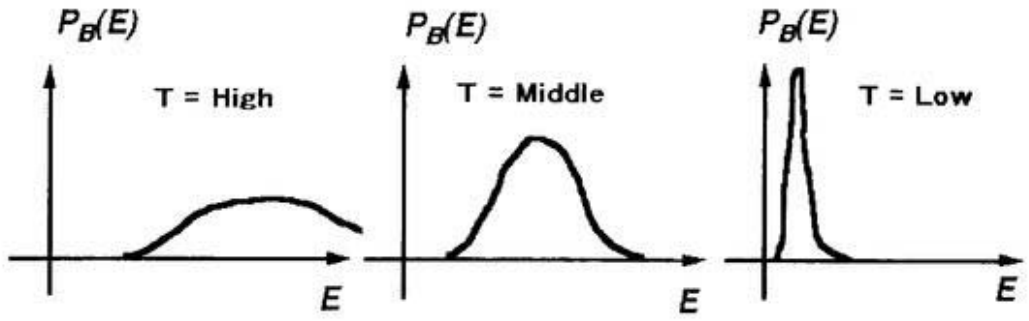


Figure 2.3: The probability distributions $P_B(E;T)$ in the canonical ensemble at high, middle, and low temperatures.

where γ is a constant which is determined by T_I, T_F, and the total number, Ns, of MC sweeps of the run. Namely, we have

$$\gamma = \frac{1}{Ns-1} \ln \frac{T_F}{T_I} \,. \tag{2.11}$$

2.2.2 Multicanonical Algorithm

While the regular Monte Carlo method is based on the Boltzmann weight which yields the canonical distribution, the multicanonical algorithm [16] is based on a non-Boltzmann weight factor. In the multicanonical ensemble the probability distribution $P_{mu}(E)$ is defined in such a way that a configuration with any energy enters with equal probability [16]:

$$P_{mu}(E) \propto n(E) w_{mu}(E) = \text{const}, \tag{2.12}$$

where $n(E)$ is the density of states. The probability distribution $P_{mu}(E)$ is shown in Fig. 2.4. Each energy has equal probability, and thus the multicanonical simulation performs a one-dimensional random walk in energy space, which enhances the escape from states of energy local minima.

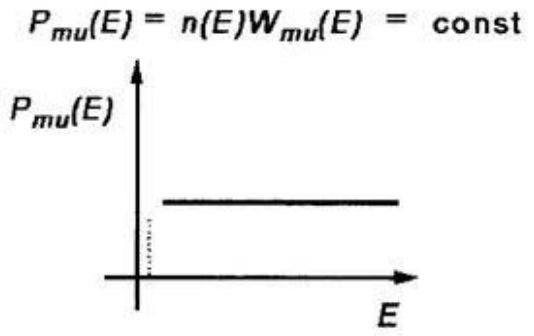


Figure 2.4: The probability distribution $P_{mu}(E)$ in the multicanonical ensemble.

From Eq. (2.12), it follows that the multicanonical weight factor $w_{mu}(E)$ should have the form

$$w_{mu}(E) \propto n^{-1}(E). \tag{2.13}$$

Since this weight factor is not *a priori* known, one has to determine it for each system by a few iterations of trial simulations. There are a few procedures to determine the multicanonical weight factor w_{mu}. In the following we give three versions of the procedures [17]-[19].

1. The multicanonical weight factor w_{mu} is defined by [17]

$$w_{mu}(E) = e^{-\beta(E)E-\alpha(E)}. \tag{2.14}$$

(1) Perform a short canonical Monte Carlo simulation at sufficiently high temperature T_0, namely $\alpha(E)$=0 and $\beta(E)$= β_0 with $\beta_0 = 1/RT_0$. The weight factor for this simulation is given by $w_{mu}^{(0)}(E)$=e$^{-\beta_0 E}$. Initialize the array $S^{(0)}(E)$ to zero, where E is discretized with bin width δE(=1 kcal/mol in the present work).

(2) Store the energy distribution obtained with the weight $w_{mu}^{(i)}(E)$ $(i \geq 0)$ as a histogram $H^{(i)}(E)$ with the same bin width as for $S^i(E)$. In the first iteration (step 1 above) determine E_{max} as the value near the mode where the histogram has its maximum (E_{max} is fixed throughout the iterations). Let E_{min} be the lowest energy obtained throughout the preceding iterations. For all $H^{(i)}(E)$ with entries greater than a certain minimum value and for the energy range $E_{min} \leq E \leq E_{max}$, update

the array $S^{(i+1)}(E)$ by

$$S^{(i+1)}(E) = \begin{cases} S^{(i)}(E) + \ln H^{(i)}(E), & \text{if } H^{(i)}(E) \geq 1, \\ S^{(i)}(E), & \text{if } H^{(i)}(E) = 0. \end{cases} \tag{2.15}$$

(3) Calculate the following multicanonical parameters $\alpha^{(i+1)}(E)$ and $\beta^{(i+1)}(E)$ from the array $S^{(i+1)}(E)$:

$$\beta^{(i+1)}(E) = \begin{cases} \beta_0, & \text{for } E \geq E_{max}, \\ \beta_0 + \frac{S^{(i+1)}(E') - S^{(i+1)}(E)}{E' - E}, & \text{for } E_{min} \leq E < E' < E_{max}, \\ \beta^{(i+1)}(E_{min}), & \text{for } E < E_{min} \end{cases} \tag{2.16}$$

and

$$\alpha^{(i+1)}(E) = \begin{cases} 0, & \text{for } E \geq E_{max}, \\ \alpha^{(i+1)}(E') + (\beta^{(i+1)}(E') - \beta^{(i+1)}(E))E', & \text{for } E < E_{max}, \end{cases} \tag{2.17}$$

where E and E' are adjacent bins in the array $S^{(i+1)}(E)$.

(4) Start a new simulation with the multicanonical weight factor defined by

$$w_{mu}^{(i+1)}(E) = e^{-\beta^{(i+1)}(E)E - \alpha^{(i+1)}(E)} . \tag{2.18}$$

(5) Iterate the last three steps until the obtained distribution $H(E)$ becomes reasonably flat in the chosen energy range.

2. The multicanonical weight factor w_{mu} is defined by [18]

$$w_{mu}(E) = e^{-S(E)} . \tag{2.19}$$

(1) Perform a short canonical Monte Carlo simulation with $S^{(0)}(E) = 0$.

(2) Store the energy distribution obtained with the weight $w_{mu}^{(i)}(E)$ $(i \geq 0)$ as a histogram $H^{(i)}(E)$.

(3) Calculate the multicanonical parameters $S^{(i+1)}(E)$ by

$$S^{(i+1)}(E) = \begin{cases} S^{(i)}(E) + \ln H^{(i)}(E), & \text{if } H^{(i)}(E) \geq 1, \\ S^{(i)}(E), & \text{if } H^{(i)}(E) = 0. \end{cases} \tag{2.20}$$

(4) Start a new simulation with the multicanonical weight factor defined by

$$w_{mu}^{(i+1)}(E) = e^{-S^{(i+1)}(E)}. \tag{2.21}$$

(5) Iterate the last three steps until the obtained distribution $H(E)$ becomes reasonably flat in the chosen energy range.

3. The multicanonical weight factor w_{mu} is defined by [19]

$$w_{mu}(E) = e^{-S(E)} = e^{-\beta(E)E+\alpha(E)}. \tag{2.22}$$

(1) Perform a short canonical Monte Carlo simulation at sufficiently high temperature, namely $\alpha(E) = 0$ and $\beta(E) = 0$.

(2) Store the energy distribution obtained with the weight $w_{mu}^{(i)}(E)$ $(i \geq 0)$ as a histogram $H^{(i)}(E)$.

(3) Calculate

$$g_0^{(i)}(E) = \frac{H^{(i)}(E')H^{(i)}(E)}{H^{(i)}(E') + H^{(i)}(E)}. \tag{2.23}$$

The normalized weights

$$\hat{g}_0^{(i)}(E) = \frac{g_0^{(i)}(E)}{g^{(i)}(E) + g_0^{(i)}(E)}, \tag{2.24}$$

$$\hat{g}^{(i)}(E) = 1 - \hat{g}_0^{(i)}(E) \tag{2.25}$$

are determined by the recursion

$$g^{(i+1)}(E) = g^{(i)}(E) + g_0^{(i)}(E), \quad g_0(E) = 0. \tag{2.26}$$

We can estimate the following multicanonical parameters $\alpha^{(i+1)}(E)$ and $\beta^{(i+1)}(E)$:

$$\beta^{(i+1)}(E) = \beta^{(i)}(E) + \hat{g}_0^{(i)}(E) \times [\ln \hat{H}^{(i)}(E') - \ln \hat{H}^{(i)}(E)]/\delta E \tag{2.27}$$

and

$$\alpha^{(i+1)}(E) = \alpha^{(i+1)}(E') + [\beta^{(i+1)}(E) - \beta^{(i+1)}(E')]E, \tag{2.28}$$

where $\hat{H}(E)$ =max$[1, H(E)]$.

(4) Start a new simulation with the multicanonical weight factor defined by

$$w_{mu}^{(i+1)}(E) = e^{-\beta^{(i+1)}(E)E+\alpha^{(i+1)}(E)}. \tag{2.29}$$

(5) Iterate the last three steps until the obtained distribution $H(E)$ becomes reasonably flat in the chosen energy range.

Once the weight factor $w_{mu}(E)$ is determined, one performs with this weight factor a long multicanonical simulation. Monitoring the energy in this simulation, one would see

that a random walk between high energy states and the ground-state configurations are realized. In this way information is collected over the whole energy range. Finally, from this simulation one can not only locate the energy global minimum but also calculate the ensemble average of any physical quantity at any temperature for a wide range of temperatures by the reweighting techniques [20] as explained in detail below.

The reweighting technique were originally developed to obtain the information of the canonical distributions at various temperature from a MC run at a single temperature [20]. Suppose this single MC run was performed at temperature $T = T_q$, the probability distribution of energy is given by

$$P_B(E;T_q) = \frac{n(E)e^{-\beta_q E}}{\int dE n(E)e^{-\beta_q E}}, \tag{2.30}$$

where $\beta_q = 1/k_B T_q$. The probability distribution at another temperature $T_r = 1/k_B\beta_r$ is then given by

$$P_B(E;T_r) = \frac{P_B(E,T_q)e^{-(\beta_r-\beta_q)E}}{\int dE P_B(E,T_q)e^{-(\beta_r-\beta_q)E}} . \tag{2.31}$$

This means that once we have the distribution at temperature T_q, we can use this distributions to calculate the canonical distributions at another temperature T_r by Eq. (2.31). However, canonical distributions are usually rather localized, and the reweighting is valid only for temperature in the vicinity of T_q. Multicanonical algorithm, however, allows one to explore all parts of the energy surface with equal probability. Thus the combination of reweighting techniques and multicanonical algorithm yields the canonical distribution over a wide range of temperatures from a single simulation run as follows:

$$P_B(E;T) = \frac{P_{mu}(E)w_{mu}^{-1}(E;T)e^{-\beta E}}{\int dE\ P_{mu}(E)w_{mu}^{-1}(E;T)e^{-\beta E}}, \tag{2.32}$$

where $P_{mu}(E)$ is the distribution of energy obtained by the multicanonical simulation. The thermodynamic average of a practical quantity A at temperature T is then obtained by

$$< A >_T = \frac{\int dE\ A(E)P_B(E;T)}{\int dE\ P_B(E;T)} . \tag{2.33}$$

Bibliography

[1] a) F. A. Momany, R. F. McGuire, A. W. Burgess, and H. A. Scheraga, *J. Phys. Chem.* **79**, 2361 (1975).

b) G. Némethy, M. S. Pottle, and H. A. Scheraga, *J. Phys. Chem.* **87**, 1883 (1983).

c) M. J. Sippl, G. Némethy, and H. A. Scheraga, *J. Phys. Chem.* **88**, 6231 (1984).

[2] H. Kawai, Y. Okamoto, M. Fukugita, T. Nakazawa, and T. Kikuchi, *Chem. Lett.* **1991**, 213.

[3] Y. Okamoto, M. Fukugita, T. Nakazawa, and H. Kawai, *Protein Eng.* **4**, 639 (1991).

[4] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, *J. Chem. Phys.* **24**, 1437 (1985).

[5] M. Irisa, T. Takahashi, F. Hirata, and T. Yanagida, *J. Mol. Liquids.* **65**, 381 (1995).

[6] M. Irisa, T. Takahashi, K. Nagayama, and F. Hirata, *Mol. Phys.* **85**, 1227 (1995).

[7] M. Irisa, *Comput. Phys. Comm.* **98**, 317 (1996).

[8] a) B. Honig and A. Nicholls, *Science* **268**, 1144 (1995).

b) H. Nakamura, *Quart. Rev. of Biophys.* **29**, 1 (1996).

[9] T. Takahashi, H. Nakamura, and A. Wada, *Biopolymers*, **32**, 897 (1992).

[10] T. Ooi, M. Oobatake, G. Némethy, H. A. Scheraga, *Proc. Natl. Acad. Sci. USA* **84**, 3086 (1987).

[11] L. Wesson and D. Eisenberg, *Protein Sci.* **1**, 227 (1992).

[12] M. Masuya and Y. Okamoto, manuscript in preparation.

[13] M. Masuya, manuscript in preparation.

[14] F. Eisenhaber, P. Lijnzaad, P. Argos, C. Sander, and M. Scharf, *J. Comp. Chem.* **16**, 273 (1995).

[15] S. Kirkpatrick, C. D. Gelatt, Jr., and M. P. Vecchi, *Science* **220**, 671 (1983).

[16] B. A. Berg and T. Neuhaus, *Phys. Lett.* **B267**, 249 (1991).

[17] Y. Okamoto and U. H. E. Hansmann, *J. Phys. Chem.* **99**, 11276 (1995).

[18] J. Lee, *Phys. Rev. Lett.* **71**, 211 (1993).

[19] B. A. Berg, *Nucl. Phys. B (Proc. Suppl.)* **63A-C**, 98 (1998).

[20] A. M. Ferrenberg and R. H. Swendsen, *Phys. Rev. Lett.* **61**, 2635 (1988).

Chapter 3

Classification of Low-Energy Conformations of Met-Enkephalin in Gas Phase and in a Model Solvent Based on the Extended Scaled Particle Theory

Ayori Mitsutake, Masayuki Irisa, Yuko Okamoto, and Fumio Hirata, "Classification of Low-Energy Conformations of Met-Enkephalin in the Gas Phase and in a Model Solvent Based on the Extended Scaled Particle Theory," *Bulletin of the Chemical Society of Japan* **72**, 1717-1729 (1999).

3.1 Introduction

The penta peptide Met-enkephalin in gas phase is one of the biomolecules most frequently studied by computer simulations.[1]-[11]. The global-minimum-energy conformation in gas phase is known [2] and the low-energy conformations have been studied in detail [4, 5, 9]. In this chapter we combine Monte Carlo simulated annealing [12] and the extended scaled particle theory (ESPT) [13, 14] to study the low-energy conformations of Met-enkephalin. The scaled particle theory [15, 16] is a method based on the statistical mechanics of liquids, which is known to give good account of the hydration free energy of a non-polar solute. It has been successfully used to estimate the solvation free energy of biomolecules with a fixed conformation [13, 14], but has never been applied to simulations of protein folding. We performed Monte Carlo simulated annealing simulations of Met-enkephalin in gas phase and in aqueous solution which is represented by the cavity-formation free energy. The simulations were repeated twenty times from randomly generated initial conformations, and the low-energy conformations were classified into groups of similar structures for both environments.

3.2 Computational Details

We employ the Monte Carlo simulated annealing (SA) for obtaining the lowest-energy conformation and the scaled particle theory for including solvent effects. For Met-enkephalin (whose amino-acid sequence is Tyr-Gly-Gly-Phe-Met), the number of degrees of freedom (namely, the number of dihedral angles to be updated) is 19. One Monte Carlo (MC) sweep consists of updating all 19 angles once with a Metropolis evaluation [17] for each update. In the present work, each SA run of Met-enkephalin consisted of 10000 MC sweeps with the initial temperature of 1000 K and the final temperature of 150 K. The temperature was lowered exponentially [18] (see the subsection 2.2.1). We made 40 SA runs altogether, starting from random initial conformations. Twenty SA runs were carried out for a simulation in gas phase (namely, with only the conformational energy term (E_P) (see Eq. (2.3)). The other twenty SA runs were for a simulation in the solvent (namely,

with both the conformational energy term (E_P) and the cavity-formation free energy term ($\mathcal{E}_{CAV}$) (see Eqs. (2.3), (2.3), and (2.6))).

For the calculation of $\mathcal{E}_{CAV}$, which is based on ESPT (see the subsection 2.1.2), we used a constant temperature of 298.15 K for simplicity. The density of the water molecules was thus set equal to 1.0 g/cm^3. To calculate the excluded volumes, all hydrogen atoms in the system were regarded as being absorbed into the united atoms. The radii of all the united solute atoms were fixed to 1.9 Å for simplicity. That of the water molecule was taken to be 1.4 Å.

3.3 Results and Discussion

3.3.1 Previous Results

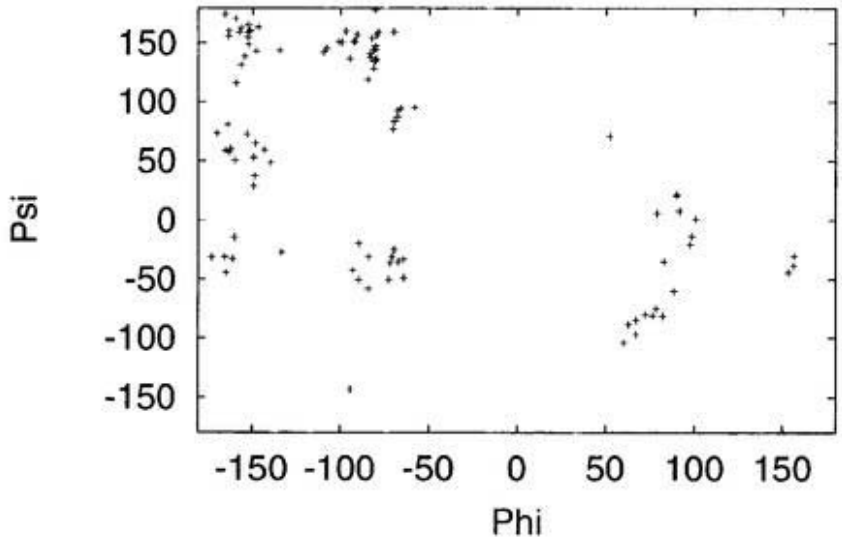
In this subsection, we summarize the results of previous works [5, 9] that classified low-energy conformations of Met-enkephalin. In Ref. [5] Monte Carlo simulated annealing [12] simulations in gas phase with Met-enkephalin were performed and the 40 obtained lowest-energy conformations were analyzed by the root-mean-square distances. These conformations were classified into four characteristic groups of similar structures (they were referred to as groups A, B, C, and D). These conformations correspond to the local-minimum energy conformations of Met-enkephalin.

On the other hand, in Ref. [9] the pattern of hydrogen bonding was used to characterize groups of low-energy structures of Met-enkephalin. However, the investigations focused on conformations with energies little above the ground state and therefore found here only structures corresponding to groups A (the lowest-energy state) and B (the second-lowest-energy state) in Ref. [5]. The conformations in group A have two hydrogen bonds between Gly-2 and Met-5 and form a type II′ β-turn involving Gly-Gly-Phe-Met. They correspond to the most stable structure of Met-enkephalin in gas phase. The conformations in group B have hydrogen bonds between Tyr-1 and Phe-4 and form a type II β-turn involving Tyr-Gly-Gly-Phe.

3.3.2 Ramachandran Plots

We first study Ramachandran plots, which are the distributions of dihedral angles (ϕ,ψ), for the lowest-energy structures obtained by the present simulations. The results in gas phase and in the solvent are shown in Fig. 3.1(a) and Fig. 3.1(b), respectively. Those for each individual residue are separately shown in Fig. 3.2. The distributions of points for the amino-acid residues, except for Gly, are limited mostly to the second and third quadrants ($\phi < 0$), while those for Gly are observed in all regions. This is in accord with the distributions obtained from a protein database, implying that our simulations were effective. From Figures 3.1 and 3.2, we also find that the distributions of the dihedral angles in gas phase are slightly more localized than those in the solvent. This seems to indicate that solvation tends to smooth out the free energy landscape of Met-enkephalin. However, the results in both cases are amazingly alike, and these Ramachandran plots cannot infer any definite conformational differences between the two cases.

(a)



(b)

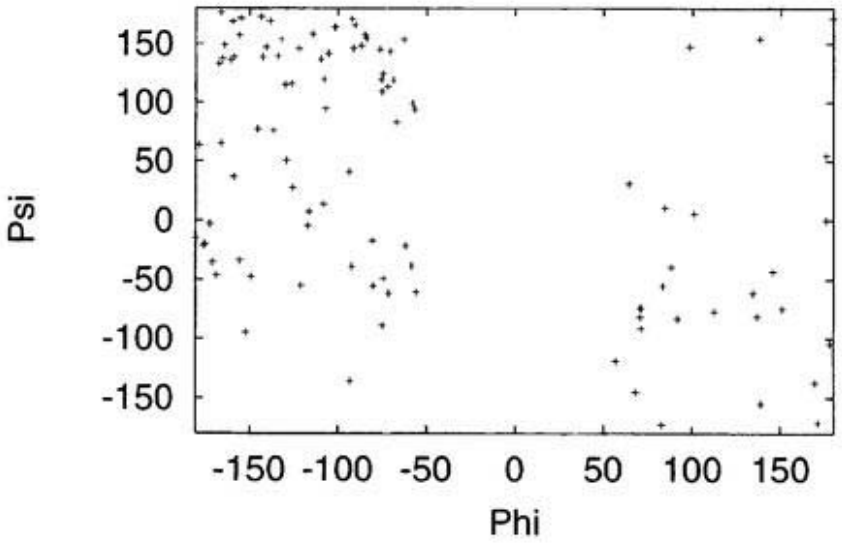
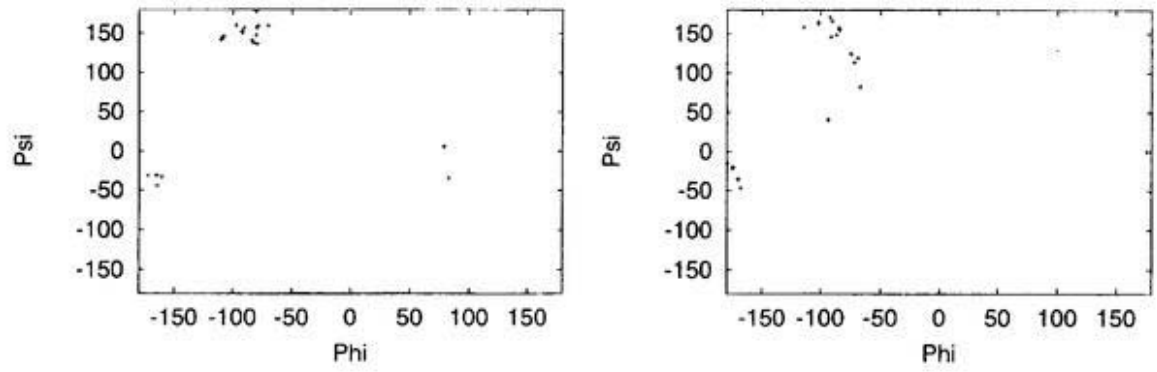


Figure 3.1: Distribution of dihedral angles (ϕ,ψ) of the lowest-energy conformations of Met-enkephalin obtained in each of the 20 Monte Carlo simulated annealing runs in gas phase (a) and in the solvent (b).

(a)



(b)

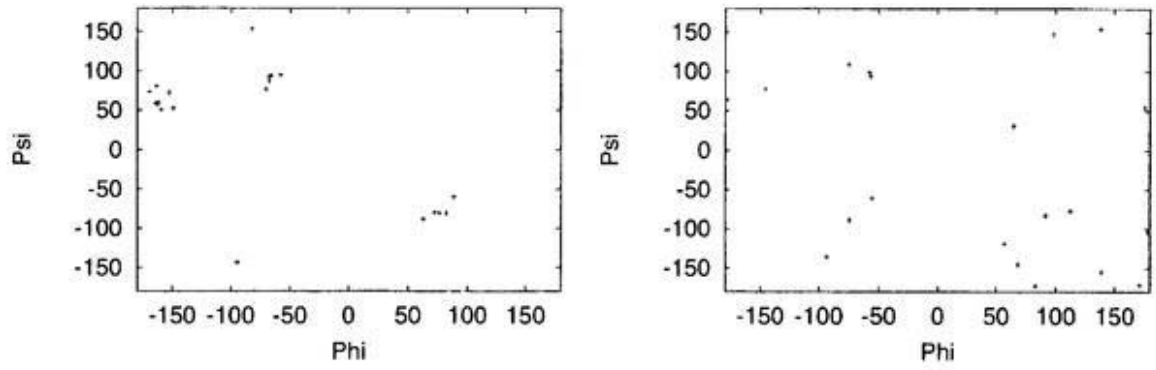
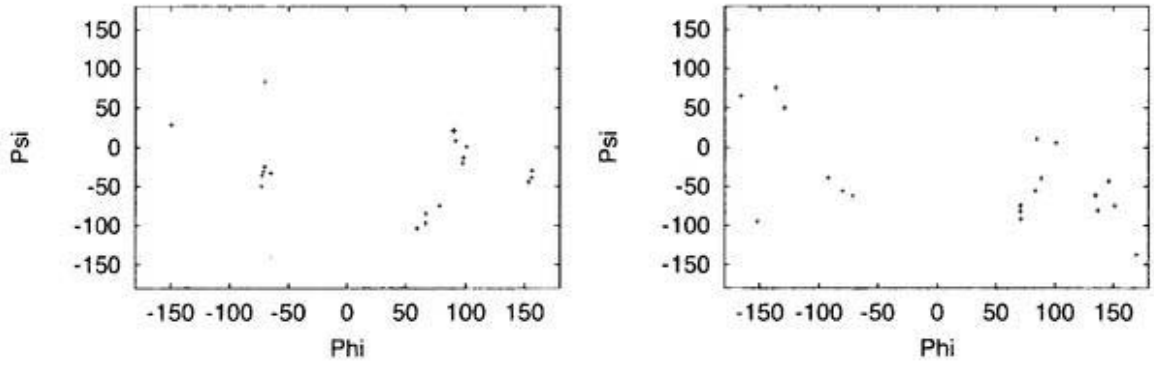
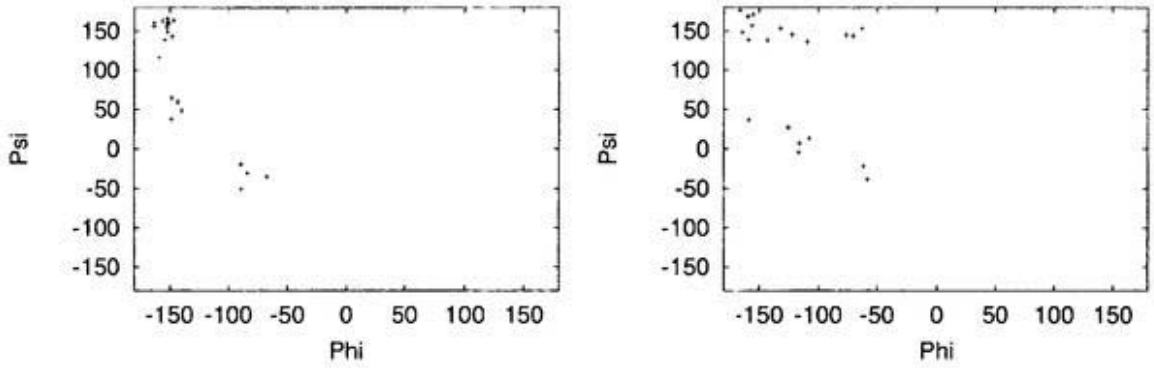


Figure 3.2: Distribution of dihedral angles of the lowest-energy conformations of Met-enkephalin for Tyr-1 (a), Gly-2 (b), Gly-3 (c), Phe-4 (d), and Met-5 (e), which were obtained by the 20 Monte Carlo simulated annealing simulations in gas phase (left-hand side) and in the solvent (right-hand side).

(c)



(d)



(e)

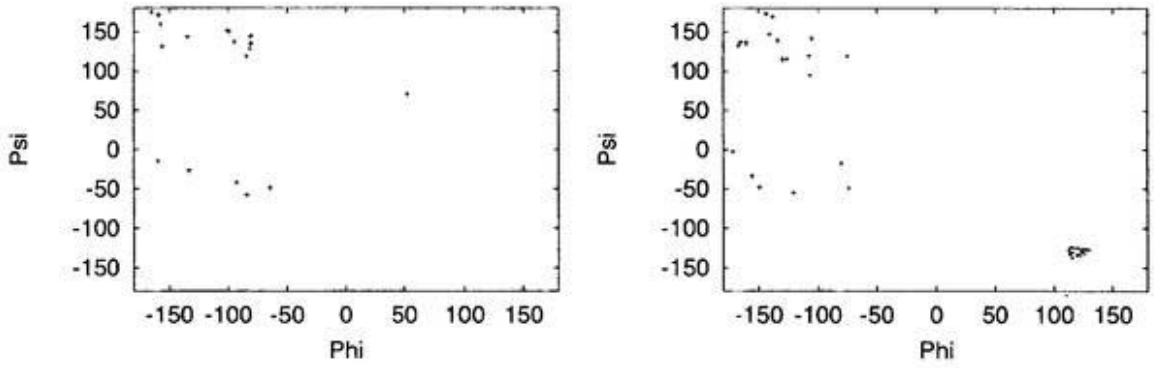


Fig. 3.2 (Continued)

3.3.3 Definition for Classification of Conformations

In order to compare the three-dimensional structures obtained by the present simulations in more detail, we calculated the root-mean-square distances (RMSD) between all pairs of the 40 lowest-energy conformations and four structures which were obtained in a previous work [19] as the characteristic conformations. We took into account only the atoms in the backbone. From these results, we classified the conformations into a few groups of similar structures. We define a group of conformations as follows [18]. Let W be the set of all conformations under consideration. A subset Y (with more than one conformation) of W is called a *group* of similar structures if it satisfies the following condition: For any conformation K in Y, there exists at least another conformation K′ in Y so that RMSD between conformations K and K′ is less than a cutoff c. (The definition depends on the value of c.) Each group corresponds to a specific tertiary structure and its small variations.

3.3.4 Detailed Results of Classification in Gas Phase

The RMSD in gas phase are listed in Table 3.1. The conformations from column 2 to column 5 (A, B, C, and D) are the lowest-energy structures obtained previously in gas phase, which represent the four groups found in Ref. [19]. The other 20 conformations (V1, V2, $\cdots$, and V20) are the lowest-energy structures obtained by the present simulations in gas phase. The numbers enclosed in the solid lines, and the dashed lines in Table 3.1 indicate that the corresponding pairs of conformations form a group of similar structures with the cutoff $c = 1.0$ Å and $c = 1.5$ Å, respectively. We thus have, respectively, six groups (six boxes in solid lines) and three groups (three boxes in dashed lines) for the two cutoff values. The numbers in bold in Table 3.1 from the second to the fifth column are the entries that are less than or equal to 1.0. These entries imply that the group which includes conformations V1, V2, and V3 has similar structures to A. Likewise, the group which includes conformations V6, V7, and V8 has similar structures to D, and that including V10 and V11 to B. Hence, we refer to these three groups as groups A_1 (including V1,V2, and V3), D_1(including V6, V7, and V8), and B_1(including V10 and

V11) (as defined with cutoff $c = 1.0$ Å). Furthermore, the underlined numbers in Table 3.1 indicate that they are more than 1.0 and less than or equal to 1.5. These entries imply that conformations V13, V14, V17, and V18 have similar structures to group B. We thus refer to the group including V13, V14, V15, and V16 as B_1' and that including V17 and V18 as B_1'' (as defined with cutoff $c = 1.0$ Å). The remaining group was named E_1 (including V19 and V20). With the cutoff $c = 1.5$ Å, it is found that conformation V4 and those of group A_1 form a single group, which we refer to as group A_2 and that conformations V9 and V12 and those of groups D_1, B_1, B_1', and B_1'' constitute a single group, which we refer to as group B_2. All these results of classifications are listed in Table 3.2.

Table 3.1 : The root-mean-square distances (Å) among pairs of the twenty lowest-energy conformations obtained by Monte Calro simulated annealing simulations in gas phase.[a]

	A	C	B	D	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15	V16	V17	V18	V19	V20
V1	**0.2**	2.3	2.7	1.7		0.3	0.3	1.1	2.0	1.9	1.9	2.0	2.3	2.5	2.6	2.6	2.7	2.7	2.7	2.7	2.8	2.6	2.8	2.8
V2	**0.2**	2.2	2.6	1.7	0.3		0.3	1.2	1.9	1.9	1.8	1.9	2.2	2.5	2.5	2.5	2.6	2.6	2.6	2.6	2.7	2.6	2.8	2.8
V3	**0.2**	2.2	2.6	1.7	0.3	0.3		1.1	1.9	1.9	1.8	1.9	2.2	2.4	2.5	2.5	2.6	2.6	2.6	2.6	2.7	2.6	2.8	2.9
V4	1.1	2.3	2.6	1.7	1.1	1.2	1.1		2.1	1.9	1.8	1.9	2.0	2.4	2.6	2.4	2.5	2.6	2.6	2.5	2.8	2.7	2.8	2.9
V5	2.0	1.7	1,7	1,8	2.0	1.9	1.9	2.1		1.7	1.7	1.6	1.8	1.6	1.7	2.1	1.9	1.9	2.0	2.0	2.0	1.9	2.7	2.8
V6	1.9	2.5	1.6	**0.9**	1.9	1.9	1.9	1.9	1.7		0.7	0.9	1.4	1.6	1.8	1.8	1.8	1.9	2.0	2.0	1.8	1.8	2.2	2.5
V7	1.8	2.5	1.8	**0.7**	1.9	1.8	1.8	1.8	1.7	0.7		0.7	1.3	1.4	1.7	1.7	1.9	2.0	1.9	1.9	2.0	1.9	2.5	2.6
V8	1.9	2.3	1.7	**0.8**	2.0	1.9	1.9	1.9	1.6	0.9	0.7		1.2	1.4	1.6	1.4	1.8	1.9	1.8	1.8	1.8	1.7	2.4	2.5
V9	2.2	2.1	1.5	1.4	2.3	2.2	2.2	2.0	1.8	1.4	1.3	1.2		1.2	1.5	1.2	1.7	1.7	1.6	1.7	1.8	1.7	2.5	2.6
V10	2.5	1.9	**1.0**	1.6	2.5	2.5	2.4	2.4	1.6	1.6	1.4	1.4	1.3		0.9	1.4	1.6	1.7	1.6	1.7	1.7	1.6	2.7	2.9
V11	2.6	2.0	**0.7**	1.8	2.6	2.5	2.5	2.6	1.7	1.8	1.7	1.6	1.5	0.9		1.4	1.4	1.5	1.4	1.5	1.5	1.5	2.5	2.5
V12	2.6	2.1	1.5	1.6	2.6	2.5	2.5	2.4	2.1	1.8	1.7	1.4	1.2	1.4	1.4		1.7	1.7	1.6	1.6	1.4	1.3	2.3	2.4
V13	2.6	2.2	1.2	1.9	2.7	2.6	2.6	2.5	1.9	1.8	1.9	1.8	1.7	1.6	1.4	1.7		0.9	1.1	1.2	1.7	1.7	2.4	2.6
V14	2.7	2.1	1.4	2.0	2.7	2.6	2.6	2.6	1.9	1.9	2.0	1.9	1.7	1.7	1.5	1.7	0.9		0.7	0.7	1.5	1.4	2.7	2.9
V15	2.6	2.1	1.6	1.9	2.7	2.6	2.6	2.6	2.0	2.0	1.9	1.8	1.6	1.6	1.4	1.6	1.1	0.7		0.3	1.7	1.6	2.8	2.9
V16	2.6	2.0	1.7	2.0	2.7	2.6	2.6	2.5	2.0	2.0	1.9	1.8	1.7	1.7	1.5	1.6	1.2	0.7	0.3		1.7	1.5	2.9	3
V17	2.8	2.1	1.3	2.0	2.8	2.7	2.7	2.8	2.0	1.8	2.0	1.8	1.8	1.7	1.5	1.4	1.7	1.5	1.7	1.7		0.4	2.1	2.3
V18	2.8	1.9	1.3	1.9	2.6	2.6	2.6	2.7	1.9	1.8	1.9	1.7	1.7	1.6	1.5	1.3	1.7	1.4	1.6	1.5	0.4		2.3	2.5
V19	2.8	3.0	2.3	2.2	2.8	2.8	2.8	2.8	2.7	2.2	2.5	2.4	2.5	2.7	2.5	2.3	2.4	2.7	2.8	2.9	2.1	2.3		0.8
V20	2.8	3.1	2.5	2.3	2.8	2.8	2.9	2.9	2.8	2.5	2.6	2.5	2.6	2.9	2.5	2.4	2.6	2.9	2.9	3.0	2.3	2.5	0.8	

[a]Conformations from column 2 to column 5 correspond to the lowest-energy conformations in groups A, B, C, and D obtained by previous simulations. The other 20 conformations (V1, V2,$\cdots$, and V20) correspond to the lowest-energy conformations obtained by the present simulations in gas phase. The numbers enclosed in solid lines and dashed lines indicate that the corresponding conformations form a group of similar structures with the cutoff $c = 1.0$ Å and $c = 1.5$ Å, respectively. The numbers in bold from the second to the fifth column are the entries that are less than or equal to 1.0. The underlined numbers indicate that they are more than 1.0 and less than or equal to 1.5.

Table 3.2: The results of classification of the lowest-energy conformations obtained by the present simulations in gas phase.

c^a (Å)	Group	n^b	Lowest-energy^c (kcal/mol)
1.0	A_1={1,2,3}	3	−11.2
	D_1={6,7,8}	3	−7.3
	B_1={10,11}	2	−9.7
	B_1'={13,14,15,16}	4	−7.2
	B_1''={17,18}	2	−7.0
	E_1={19,20}	2	−8.2
1.5	A_2={4} ∪ A_1	4	−11.2
	B_2={9,12} ∪ D_1 ∪ B_1 ∪ B_1' ∪ B_1''	13	−9.7
	E_2=E_1	2	−8.2

[a] c is the cutoff value.
[b] n is the number of structures in each group.
[c] In the fourth column we list the lowest-energy value in each group.

We now summarize the main features in the results of the classification. The lowest-energy conformation obtained throughout the present simulations (global-minimum-energy structure) is essentially identical with that obtained in previous simulations [19, 2] and belongs to group A_1. Groups A_2 and B_2 are dominant in gas phase. Group D_1 represents an intermediate structure between group A_1 and group B_1.

After the classification of the conformations into groups of similar structures is finished, we now examine the structural characterizations of each group in detail. Here, we only consider conformations of the six groups (A_1, D_1, B_1, B_1', B_1'', and E_1 in Table 3.2) obtained with the cutoff c = 1.0 Å. In Fig. 3.3 we show the lowest-energy conformations (left-hand side) and the superposed backbone structures of all conformations (right-hand side) in each group.

As can be seen in Fig. 3.3(a), the structures in group A_1 have two hydrogen bonds between the amide nitrogen of Gly-2 and the carbonyl oxygen of Met-5 backbone and between the carbonyl oxygen of Gly-2 and the amide nitrogen of Met-5 backbone. This structure is actually a type II′ β-turn involving the residues Gly-Gly-Phe-Met. Moreover, OH in the Tyr-1 side chain is hydrogen-bonded to the carbonyl oxygen of the Gly-3 backbone. It was found that the side-chain structures of Tyr-1 and Met-5 are very stable,

while that of Phe-4 is slightly more flexible. The structure of group A_1 is very stable, because of these intrachain hydrogen bonds. For instance, the conformations in group A_1 are more similar to each other than those in the other groups (RMSD are as small as 0.3 Å in Table 3.1).

As can be seen in Fig. 3.3(b), the structures in group B_1 have two hydrogen bonds between the amide nitrogen of Tyr-1 and the carbonyl oxygen of Phe-4 and between the carbonyl oxygen of Tyr-1 and the amide nitrogen of Phe-4 backbone. This structure is actually a type II β-turn involving the residues Tyr-Gly-Gly-Phe. These structures are also stable, because they have two intrachain hydrogen bonds and this group has the second-lowest energy among the six groups (group A_1 has the global-minimum energy).

The main-chain structures of groups B_1' and B_1'' are similar to those of group B_1, as can be seen in Figs. 3.3(c) and 3.3(d). The structures in these groups are slightly distorted from those in group B_1 in the sense that they have only one of the two intrachain hydrogen bonds of group B_1 formed. Group B_1' has only one hydrogen bond between the amide nitrogen of Tyr-1 and the carbonyl oxygen of Phe-4, and B_1'' has a hydrogen bond between the carbonyl oxygen of Tyr-1 and the amide nitrogen of Phe-4. These structures have slightly higher energies than those of group B_1. Also note that the side-chain structures are more deviated compared to the backbone.

As can be seen in Fig. 3.3(e), group D_1 has intermediate structures between group A_1 and B_1. This group has a hydrogen bond between the carbonyl oxygen of Gly-2 and the amide nitrogen of Phe-4, and forms a γ turn. As in the case for group A_1, OH in the Tyr-1 side chain is hydrogen-bonded to the carbonyl oxygen of the Gly-3 backbone, and the side-chain structure of Tyr-1 is stable. Group D_1 has a considerably higher energy than groups A_1 and B_1.

In group E_1 (Fig. 3.3(f)), the main chain is more extended than that of the other groups. There is no intrachain hydrogen bond in the backbone. There is a hydrogen bond between OH of the Tyr-1 side chain and the carbonyl oxygen of the Phe-4 backbone, which seems to stabilize the structure.

In summary, we found that groups A and B are dominant and stable in gas phase (they correspond to the global-minimum-energy state and the second-lowest-energy state,

respectively). These structures are stabilized by two intrachain backbone hydrogen bonds. As is shown in Tables 3.1 and 3.2, however, the number of obtained structures of group B (including B_1' and B_1'') is larger than that of group A.

Figure 3.3: The structural characterization of groups A_1, D_1, B_1, B_1', B_1'', and E_1 (from Table 3.2) obtained by Monte Carlo simulated annealing simulations in gas phase. The left-hand side is the lowest-energy conformation in each group, and the right-hand side is the superposition of the backbone structures of all the conformations in each group.
(a) Group A_1 (Conformation V1 in Table 3.1).
(b) Group B_1 (Conformation V10 in Table 3.1).
(c) Group B_1' (Conformation V13 in Table 3.1).
(d) Group B_1'' (Conformation V17 in Table 3.1).
(e) Group D_1 (Conformation V6 in Table 3.1).
(f) Group E_1 (Conformation V19 in Table 3.1).
We use a simplified notation such as O2 and N5, which stand for the carbonyl oxygen of the Gly-2 backbone and the amide nitrogen of the Met-5 backbone, respectively.

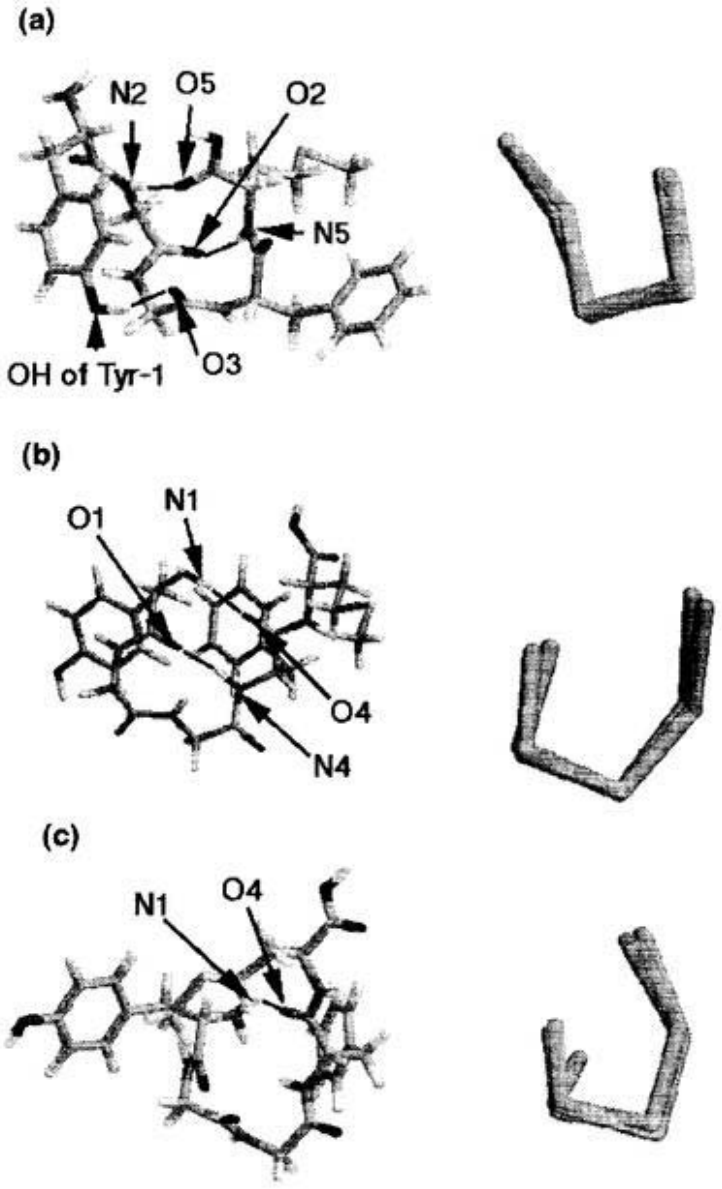


Fig. 3.3 (Continued)

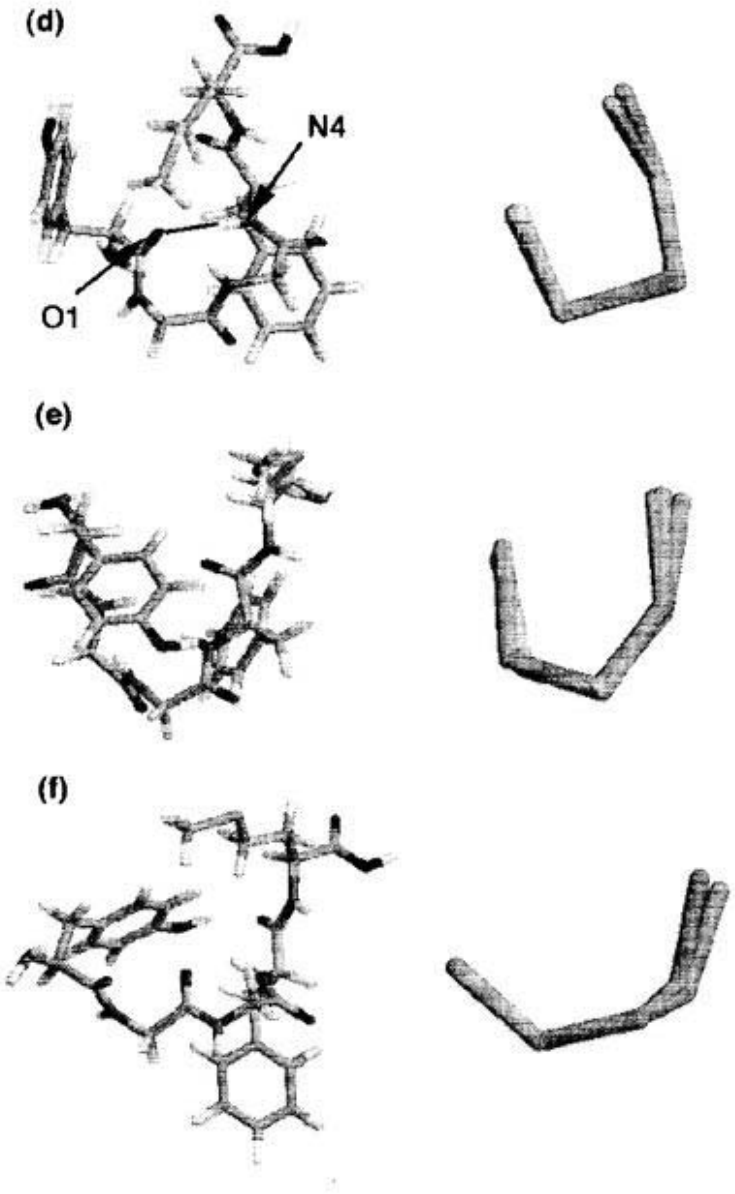


Fig. 3.3 (Continued)

3.3.5 Detailed Results of Classification in Model Solvent

We now present the results of simulations in the solvent. In Table 3.3, the RMSD in the solvent are listed. The conformations from column 2 to column 7 correspond to the lowest-energy conformations in groups A_1, D_1, B_1, B_1', B_1'', and E_1 obtained by the present simulations in gas phase, and actually are V1, V10, V13, V17, V6, and V19 in Table 3.1, respectively. The other 20 conformations (S1, S2, $\cdots$, and S20) are the lowest-energy conformations obtained by the present simulations in the solvent. The numbers enclosed in solid lines and dashed lines in Table 3.3 indicate that the corresponding conformations form a group of similar structures with the cutoff $c = 1.0$ Å and $c = 1.5$ Å, respectively. We thus have three groups (boxes in solid line and dashed line) for the two cutoff values. The numbers in bold in Table 3.3 from the second to the seventh column are those entries that are less than or equal to 1.0. These entries imply that conformations S4 and S5 have similar structures to B_1. Hence, we refer to the group including S4, S5, and S6 as group B_3 (as defined with cutoff $c = 1.0$ Å). Furthermore, the underlined numbers in Table 3.3 indicate that they are more than 1.0 and less than or equal to 1.5. These entries imply that the group which includes conformations S1 and S2 have intermediate structures between B_1 and D_1. We refer to the group including S1 and S2 as group B_3' (as defined with cutoff $c = 1.0$ Å). The remaining group including S9 and S10 was named F_3. As for the classifications of groups having similar structures with the cutoff $c = 1.5$ Å (boxes in dashed lines), it is found that conformations S3, S7, S8, S11, and S12 and groups B_3', B_3, and F_3 form a single group, which we refer to as group B_4, and that conformations S14, S15, and S16 constitute a single group, which we refer to as group G_4. The remaining group includes S17, S18, and S19, which we refer to as group E_4, because two of the entries in this group are less than 1.5 Å from group E_1. We thus find that for $c = 1.5$ Å, there are three groups (B_4, G_4, and E_4). All of these results of classifications are summarized in Table 3.4. We remark that the lowest-energy conformation obtained throughout the present simulations (global-minimum-energy structure) belongs to group B_3' (Conformation S1). In the case of the cutoff $c = 1.5$ Å, this structure belongs to the main group B_4.

After the classification of conformations into groups of similar structures is finished, we

Table 3.3: The root-mean-square distances (Å) among pairs of the twenty lowest-energy conformations obtained by Monte Calro simulated annealing simulations in solvent.[a]

	A_1	D_1	B_1	B_1'	B_1''	E_1	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20
TOT	36.4	38	38.6	39	37.5	38.5	23.7	26.5	29.8	35.9	31.7	26.1	25.7	31	33.5	33.8	33.9	37.9	35.3	29.2	31.2	33.4	30.6	27.5	33.5	33.3
SOL	47.6	45.3	48.3	46.2	44.5	46.8	24.6	27.9	26.2	40.3	40	27.9	24.4	39.6	40.7	41.1	32.4	31	37.3	32.3	25.4	27.2	26.2	22.5	27.6	31.6
GAS	-11.2	-7.3	-9.7	-7.2	-7	-8.3	-0.9	-1.4	3.6	-4.4	-8.3	-1.8	1.3	-8.6	-7.2	-7.3	1.5	6.9	-2	-3.1	5.8	6.2	4.4	5	5.9	1.7
S1	2.4	1.1	1.3	1.7	1.6	2.4		0.8	1.4	1.4	1.1	1.9	1.5	1.5	1.8	2.0	2.6	2.4	2.0	2.6	3.0	2.7	3.0	2.7	3.0	2.5
S2	2.4	1.3	1.3	1.5	1.7	2.5	0.8		1.2	1.4	1.3	1.9	1.4	1.3	1.8	2.0	2.6	2.4	1.9	2.7	3.0	2.7	3.0	2.7	2.9	2.5
S3	2.3	1.6	1.3	1.9	2.0	2.5	1.4	1.2		1.3	1.4	1.7	1.8	1.1	1.7	1.8	2.4	2.3	2.3	2.9	2.9	2.7	3.0	2.9	2.9	2.5
S4	2.6	1.5	1	1.9	1.7	2.2	1.4	1.4	1.3		1.0	1.0	1.8	1.5	1.6	1.6	2.4	2.0	2.1	3.0	2.9	2.8	2.9	2.8	2.8	2.3
S5	2.7	1.6	0.9	1.5	1.4	2.4	1.1	1.3	1.4	1.0		1.5	1.5	1.8	1.5	1.8	2.6	2.4	2.1	2.6	3.0	2.7	3.0	2.7	2.8	2.3
S6	2.6	1.8	1.6	2.2	1.9	1.9	1.9	1.9	1.7	1.0	1.5		2.2	1.6	1.8	1.6	2.0	1.8	1.8	2.8	2.3	2.4	2.4	2.5	2.3	2
S7	2.5	1.6	1.5	0.9	1.7	2.8	1.5	1.4	1.8	1.8	1.5	2.2		2.0	1.8	2.0	3.1	2.8	2.1	2.4	3.1	2.8	3.0	2.4	3.0	2.4
S8	2.0	1.4	1.7	2.1	2.2	2.1	1.5	1.3	1.1	1.5	1.8	1.6	2.0		1.8	1.7	2.2	1.9	1.8	2.7	2.5	2.4	2.6	2.6	2.6	2.3
S9	2.3	1.7	1.7	1.8	1.9	2.2	1.8	1.8	1.7	1.6	1.5	1.8	1.8	1.8		0.8	2.2	1.9	2.0	2.6	2.6	2.5	2.5	2.4	2.4	2.1
S10	2.2	1.8	2.0	2.1	2.1	2.1	2.0	2.0	1.8	1.6	1.8	1.6	2.0	1.7	0.8		1.9	1.5	2.1	2.7	2.4	2.3	2.3	2.4	2.1	1.9
S11	2.9	2.4	2.9	3.1	2.6	1.7	2.6	2.6	2.4	2.4	2.6	2.0	3.1	2.2	2.2	1.9		1.4	2.1	3.0	1.9	2.1	2.0	2.5	2.0	2.4
S12	2.7	2.1	2.6	2.8	2.5	1.8	2.4	2.4	2.3	2.0	2.4	1.8	2.8	1.9	1.9	1.5	1.4		2.2	3.1	2.4	2.5	2.0	2.5	2.2	2.6
S13	2.4	1.8	2.2	1.9	1.8	1.7	2.0	1.9	2.3	2.1	2.1	1.8	2.1	1.8	2.0	2.1	2.1	2.2		2.0	2.2	1.9	2.0	1.7	2.1	2
S14	2.0	2.5	2.7	2.4	2.2	2.6	2.6	2.7	2.9	3.0	2.6	2.8	2.4	2.7	2.6	2.7	3.0	3.1	2.0		2.1	1.4	2.3	1.7	2.5	1.9
S15	2.4	2.8	3.1	3.2	2.7	1.9	3.0	3.0	2.9	2.9	3.0	2.3	3.1	2.5	2.6	2.4	1.9	2.4	2.2	2.1		1.4	1.8	2.1	1.8	1.8
S16	2.2	2.5	2.8	2.8	2.4	1.9	2.7	2.7	2.7	2.8	2.7	2.4	2.8	2.4	2.5	2.3	2.1	2.5	1.9	1.4	1.4		1.7	1.8	2.0	1.6
S17	2.7	2.8	3.2	3.0	2.6	1.4	3.0	3.0	3.0	2.9	3.0	2.4	3.0	2.6	2.5	2.3	2.0	2.0	2.0	2.3	1.8	1.7		1.3	1.1	2.2
S18	2.6	2.6	2.9	2.3	2.1	1.7	2.7	2.7	2.9	2.8	2.7	2.5	2.4	2.6	2.4	2.4	2.5	2.5	1.7	1.7	2.1	1.8	1.3		1.4	2.1
S19	2.9	2.8	3.1	2.9	2.4	1.4	3.0	2.9	2.9	2.8	2.8	2.3	3.0	2.6	2.4	2.1	2.0	2.2	2.1	2.5	1.8	2.0	1.1	1.4		2.1
S20	1.9	2.3	2.4	2.4	2.5	2.3	2.5	2.5	2.5	2.3	2.3	2.0	2.4	2.3	2.1	1.9	2.4	2.6	2.0	1.9	1.8	1.6	2.2	2.1	2.1	

[a]Conformations from column 2 to column 7 correspond to the lowest-energy conformations in group A_1, D_1, B_1, B_1', B_1'', and E_1 obtained by the present simulations in gas phase and actually are V1, V10, V13, V17, V6, and V19 in Table 1, respectively. The other 20 conformations (S1, S2, $\cdots$, and S20) correspond to the lowest-energy conformations obtained by the present simulations in solvent. The numbers enclosed in solid lines and dashed lines indicate that the corresponding conformations form a group of similar structures with the cutoff $c = 1.0$ Å and $c = 1.5$ Å, respectively. The numbers in bold from the second to the seventh column are the entries that are less than or equal to 1.0. The underlined numbers indicate that they are more than 1.0 and less than or equal to 1.5.

Table 3.4: The results of classification of the lowest-energy conformations obtained by the present simulations in solvent.

c^a(Å)	Group	n^b	Lowest-energy^c (kcal/mol)
1.0	B'_3={1,2}	2	23.7
	B_3={4,5,6}	3	26.1
	F_3={9,10}	2	33.5
1.5	B_4={3,7,8,11,12,13} $\cup$ B'_3 $\cup$ B_3 $\cup$ F_3	13	23.7
	G_4={14,15,16}	3	29.2
	E_4={17,18,19}	3	27.5

[a] c is the cutoff value.
[b] n is the number of structures in each group.
[c] In the fourth column we list the lowest-energy value in each group.

now examine the structural characterizations of each group in detail. In Fig. 3.4 we show the lowest-energy conformations (left-hand side) and the superposed backbone structures of all conformations (right-hand side) in each group. As can been seen in Fig. 3.4(a), the structures in group B'_3 have a hydrogen bond between the amide nitrogen of Tyr-1 and the carbonyl oxygen of the Phe-4 backbone. Moreover, the side-chain of Tyr-1 is close to that of Met-5. As can be seen in Fig. 3.4(b), the structures in group B_3 have two hydrogen bonds between the amide nitrogen of Tyr-1 and the carbonyl oxygen of Phe-4 and between the carbonyl oxygen of Tyr-1 and the amide nitrogen of Phe-4. The structures of group B_3 actually form a type II β-turn involving the residues Tyr-Gly-Gly-Phe. These structures are very similar to those in group B in gas phase. The structures in group F_3 (Fig. 3.4(c)) do not have any intrachain hydrogen bonds. However, the shape of the main chain is similar to that of group B in gas phase.

In groups G_4 and E_4 (Figs. 3.4(d) and 3.4(e)), obtained with the cutoff c = 1.5 Å, the main chain is more extended than that of the other groups. There is no intrachain hydrogen bond in the backbone. However, the structures in these groups form a circular shape in which the main chain is extended, but the side chain of Tyr-1 is close to that of Met-5.

We now summarize the main features in the results of classification. Two characteristics of conformations can be obtained from the results of the present simulations in

solvent. One is that conformations tend to have the backbone of the structures in group B obtained in gas phase. (The structures in group B have a hydrogen bond between the amide nitrogen of Tyr-1 and the carbonyl oxygen of Phe-4 backbone.) The other is that the side chain of Tyr-1 tends to be close to that of Met-5, forming a circular shape as a whole, while the backbone is extended. The main-chain conformations are similar to those of group E_1 obtained in gas phase. These facts imply that the backbone structures obtained by simulations in the solvent do not differ significantly from those in gas phase. The side-chain structures, however, can be quite different. For instance, the lowest-energy conformation in group B_1 (Fig. 3.3(b)) and that in group B_3' (Fig. 3.4(a)) have very similar backbone structures (RMSD=1.3 Å) but completely different side-chain orientations. Also note that the global-minimum-energy structure in gas phase (group A_2) has a rather higher energy in solution, and is not found by the present simulations including a solvent.

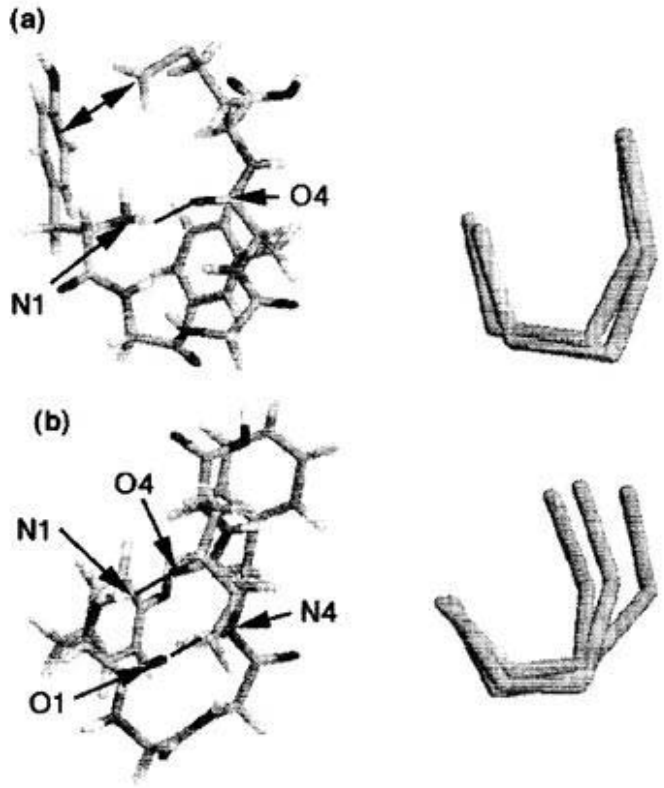


Figure 3.4: The structural characterization of groups B'_3, B_3, F_3, G_4, and E_4 (from Table 3.4) obtained by Monte Carlo simulated annealing simulations in the solvent. The left-hand side is the lowest-energy conformation in each group, and the right-hand side is the superposition of the backbone structures of all the conformations in each group.
(a) Group B'_3 (Conformation S1 in Table 3.4).
(b) Group B_3 (Conformation S6 in Table 3.4).
(c) Group F_3 (Conformation S5 in Table 3.4).
(d) Group G_4 (Conformation S14 in Table 3.4).
(e) Group E_4 (Conformation S18 in Table 3.4).
We use a simplified notation such as O2 and N5, which stand for the carbonyl oxygen of the Gly-2 backbone and the amide nitrogen of the Met-5 backbone, respectively.

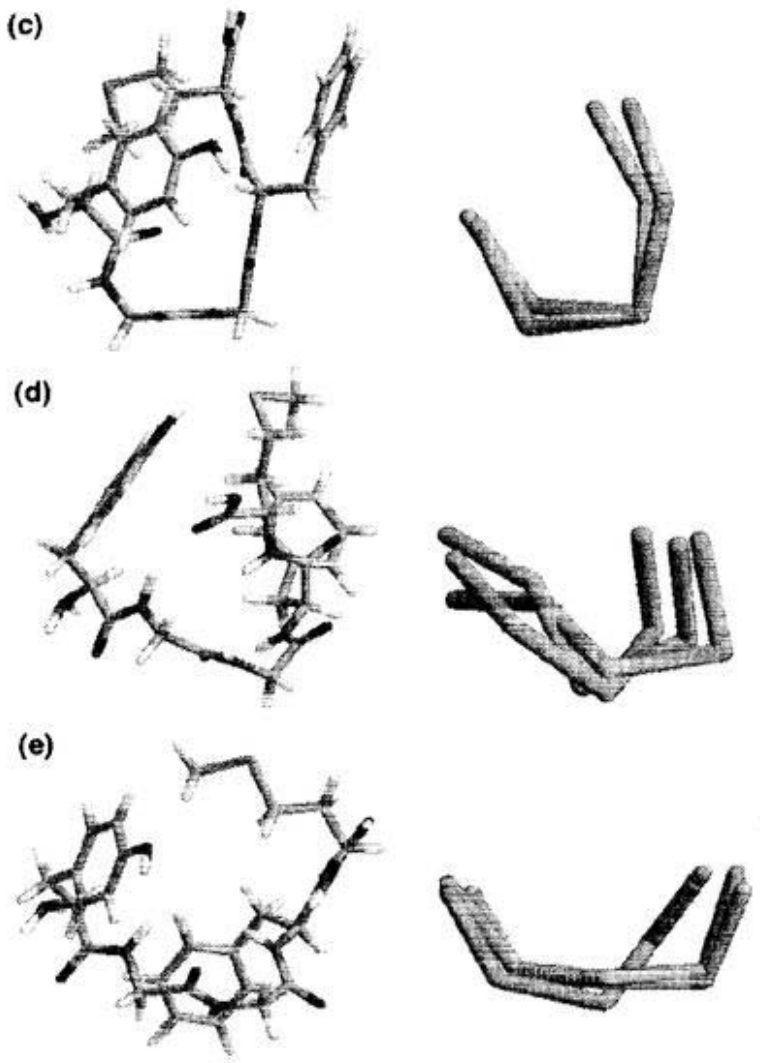


Fig. 3.4 (Continued)

3.3.6 Characteristics of the Lowest-Energy Conformations Obtained in Gas Phase and in Model Solvent

The total energy (E_{TOT}) used in the present simulations was

$$E_{TOT} = E_P \tag{3.1}$$

for gas-phase case and

$$E_{TOT} = E_P + \mathcal{E}_{CAV} \tag{3.2}$$

for the in-solvent case, where E_P is the conformational energy (see Eq. (2.3)) and $\mathcal{E}_{CAV}$ is the cavity-formation term (see Eqs. (2.5) and (2.6)). In order to see how stable the low-energy conformations in gas phase are when they are placed in the solvent, we calculated the cavity-formation term ($\mathcal{E}_{CAV}$) for six conformations (A_1, D_1, B_1, B_1', B_1'', and E_1) and list the values in Table 3.3 (see the third row in the table). As can be seen from Table 3.3, these cavity-formation terms are significantly larger than those of conformations (S1, S2, $\cdots$, and S20) obtained by simulations in the solvent, suggesting that these gas-phase conformations are unstable in the solvent. Note that the entries in Table 3.3 are the root-mean-square distances of the backbone structure only. The fact that some of the values between the conformations in gas phase and those in the solvent are small (see the entries in bold face and underlined numbers) seems to imply that when the low-energy conformations in gas phase are placed in the solvent, they rearrange the side-chain orientations so that the backbone structures remain stable. An exception is the conformation in group A_1. The side-chain structures (especially that of Tyr-1) are also constrained for these conformations so that they cannot be rearranged to find a stable conformation in the solvent. The backbone of the lowest-energy structure in the solvent has a shape similar to that of the second lowest-energy structure in gas phase; the global-minimum-energy conformation in gas phase was not found in the simulations in the solvent. Furthermore, as can be seen from Tables 2 and 4, the energy difference between the lowest-energy group and the second lowest-energy group is 1.5 kcal/mol in gas phase (A_2 versus B_2) and 3.8 kcal/mol and in the solvent (B_4 versus H_4), respectively. Therefore, there is only one dominant low-energy structure in the solvent, while there are two dominant ones in gas phase. It seems that the solvent effects decrease the number of low-energy local minima.

We now focus our attention on the conformations obtained in the solvent which have a low cavity-formation free energy ($\mathcal{E}_{CAV}$). These are conformations S1, S2, S3, S6, S7, S15, S16, S17, S18, and S19 (see Table 3.3). By examining the structures in detail, we found two characteristics. One is that the conformations with a hydrogen bond between the amide nitrogen of Tyr-1 and the carbonyl oxygen of Phe-4 backbone have low cavity-formation free energies. These conformations are S1, S2, S3, and S7, which are shown in Fig. 3.5(a). The other is that the conformations have low cavity-formation free energies when they form a circular shape in which the N-terminal side chain (Tyr-1) is close to the C-terminal one (Met-5) and the main chain is extended. These conformations are S6, S15, S16, S17, S18, and S19, which are shown in Fig. 3.5(b). In Fig. 3.6, cavity-formation free energy of the lowest-energy conformations obtained by each simulation in the solvent as a function of the accessible surface area is shown. The accessible surface areas of the obtained conformations were calculated by the code developed in Ref. [20]. It is found that there are two parts in Fig. 3.6. One part (Part 1) corresponds to the structures that have high cavity-formation free energies and small accessible surface areas. These conformations are similar to those of group B in gas phase, which have two intrachain hydrogen bonds and have low conformational energies. In this case, the conformational energies are responsible for the conformational stability in the solvent. On the other hand, the second part (Part 2) in Fig. 3.6 corresponds to the structures that have low cavity-formation free energies and large accessible surface areas. These structures have either circular shapes in which the main chain is extended, but the N-terminal side chain (Tyr-1) is close to the C-terminal one (Met-5), or have a hydrogen bond between the amide nitrogen of Tyr-1 and the carbonyl oxygen of Phe-4 (one hydrogen bond of group B are broken); also, these conformations have low cavity-formation free energies. In this case, the cavity-formation free energy is responsible for the conformational stability in the solvent.

(a)

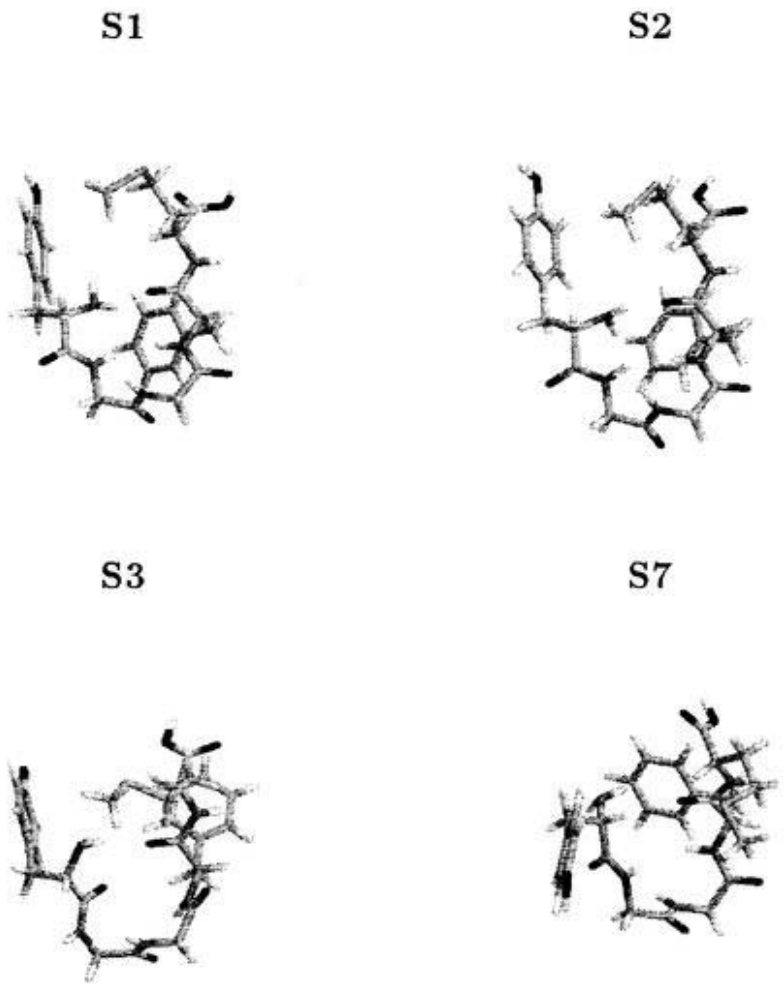
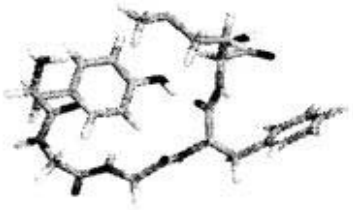


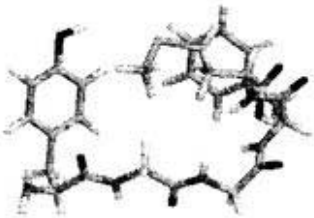
Figure 3.5: Two kinds of conformations with low cavity-formation free energy: Conformations S1, S2, S3, and S7 (a), and S6 and from S15 to S19 (b).
Structures in (a) have a hydrogen bond between the amide nitrogen of Tyr-1 and the carbonyl oxygen of Phe-4 backbone. Structures in (b) have circular shapes in which the main chain is extended and the N-terminal side-chain (Tyr-1) is very close to the C-terminal one (Met-5).

(b)

S6

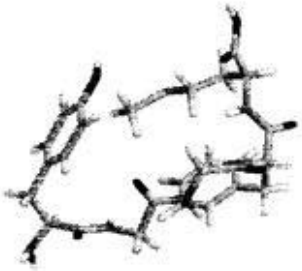


S15

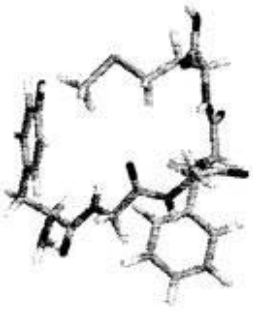


S16

S17



S18



S19

Fig. 3.5 (Continued)

3.3.7 Relation between the Cavity-Formation Free Energy and the Accessible Surface Area

We now consider in detail the relations among the solvent-accessible surface area, the solvent-excluded volume, and the cavity-formation free energy. We performed three SA runs of 10000 MC sweeps in the model solvent with the initial temperature of 1000 K and the final temperature of 150 K. One of the simulation runs, which we refer to as Run 1, was made so that it reproduces the one that gave the lowest-energy conformation in Part 1 of Fig. 3.6. Another run (Run 2) reproduces the lowest-energy conformation in Part 2 of Fig. 3.6. The third run (Run 3) was newly made. One conformation at each temperature with an increment of 1 degree (T=150, 151, 152, $\cdots$, and 1000 K) was chosen for the analyses (total of 851 conformations for each run).

In Fig. 3.7, we show the solvent-excluded volume of these conformations as a function of their accessible surface area. The excluded volumes of these conformations were calculated by the code developed in Ref. [20]. It was found that the excluded volume is almost proportional to the accessible surface area.

In Fig. 3.8, the cavity-formation free energies of the conformations as a function of the accessible surface area are shown. In Fig. 3.8(a), the data from all 851 $\times$ 3 conformations are shown. The data in Fig. 3.8(b) and Fig. 3.8(c) are from the conformations which have $E_{TOT} < 40$ kcal/mol in Run 1 and Run 2, respectively (the results of Run 3 are omitted because they turned out to be similar to those of Run 2). From Fig. 3.8, it is found that the low-energy conformations obtained in these simulations remain in Part 1 or Part 2 of Fig. 3.6. Fig. 3.8(b) implies that the cavity-formation free energy of the lower-energy conformations ($E_{TOT} < 40$ kcal/mol) in Run 1 is proportional to the accessible surface area. On the other hand, Fig. 3.8(c) suggests that the cavity-formation free energy of low-energy conformations ($E_{TOT} < 40$ kcal/mol) in Run 2 has no correlation with the accessible surface area. These results show that cavity-formation free energy is not necessarily proportional to accessible surface area for a small peptide, such as Met-enkephalin, and depends on the microscopic structures of the peptide. This conclusion is

in accord with the results of simulations of Met-enkephalin with RISM theory [21].

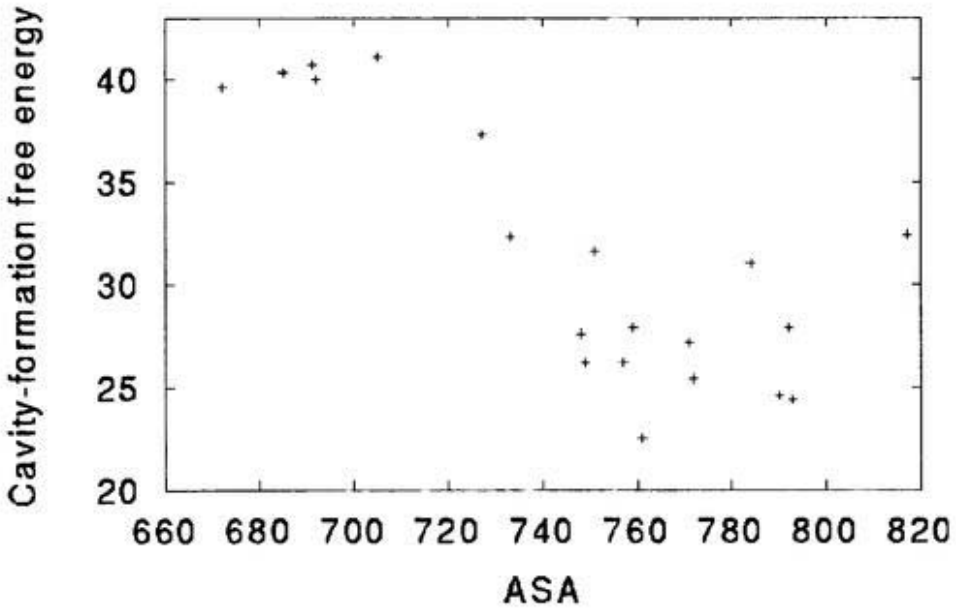


Figure 3.6: Cavity-formation free energy (kcal/mol) of conformations as a function of solvent-accessible surface area (Å^2). The conformations are the lowest-energy conformations obtained in each of the 20 Monte Carlo simulated annealing runs in the solvent. It is found that there are two parts. One part (Part 1) corresponds to the structures that have high cavity-formation free energies and small accessible surface areas. These conformations are similar to those of group B in gas phase, which have two intrachain hydrogen bonds. The second part (Part 2) corresponds to the structures that have low cavity-formation free energies and large accessible surface areas. These structures have either circular shapes in which the main chain is extended but the N-terminal side chain (Tyr-1) is close to the C-terminal one (Met-5) or have a hydrogen bond between the amide nitrogen of Tyr-1 and the carbonyl oxygen of Phe-4 (one hydrogen bond of group B are broken).

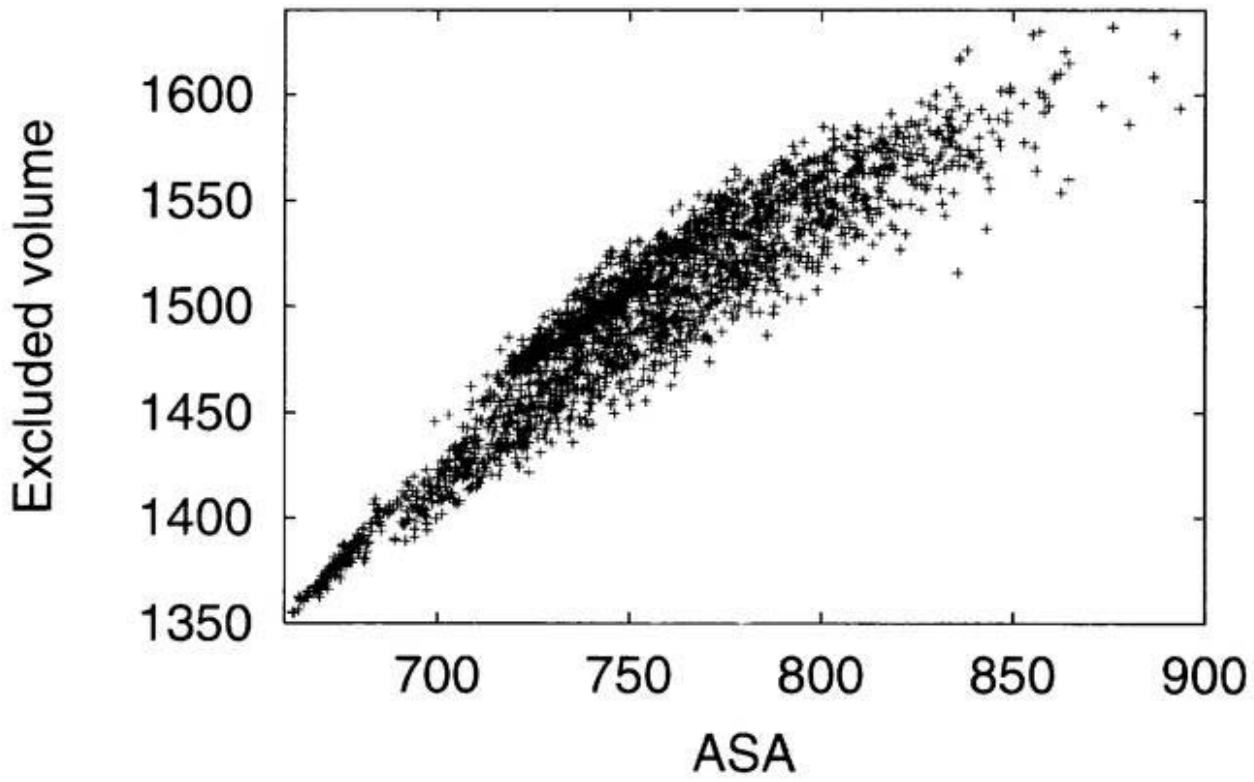


Figure 3.7: The solvent-excluded volume (Å^3) of conformations obtained by three SA runs as a function of their accessible surface area (Å^2). One of the simulation runs, which we refer to as Run 1, was made so that it reproduces the one that gave the lowest-energy conformation in Part 1 of Fig. 3.6. Another run (Run 2) reproduces the lowest-energy conformation in Part 2 of Fig. 3.6. The third run (Run 3) was newly made. One conformation at each temperature with the increment of 1 degree (T=150, 151, 152, $\cdots$, and 1000 K) was chosen for the analysis (total of 851 conformations for each run).

(a)

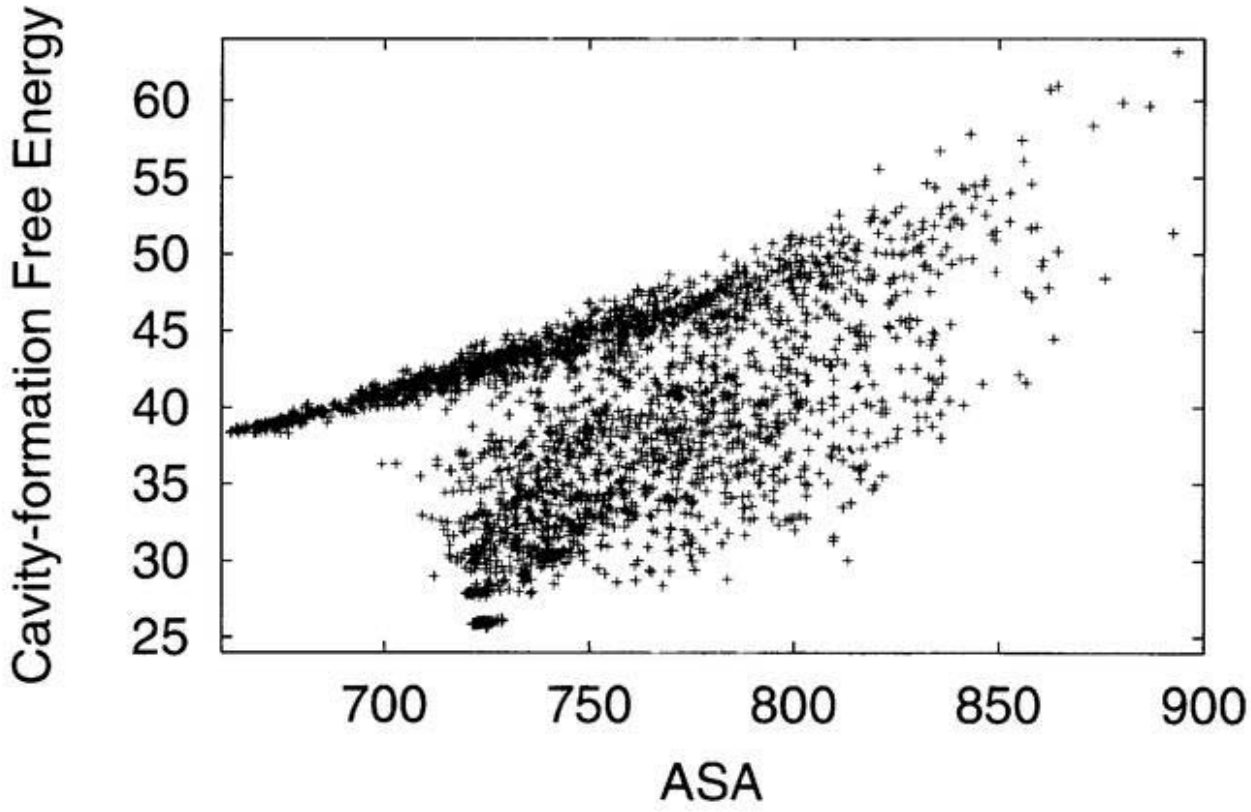
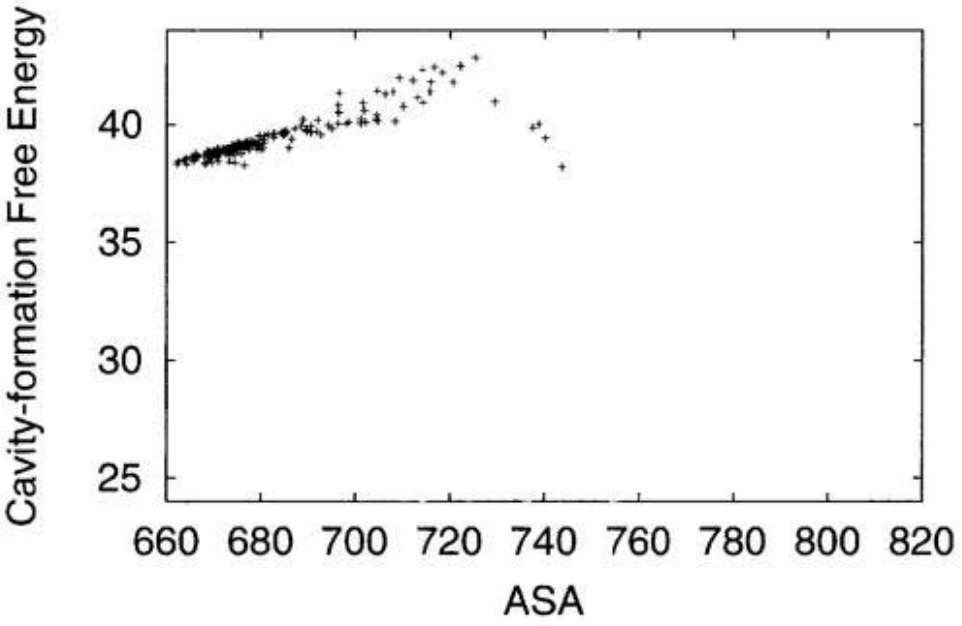


Figure 3.8: Cavity-formation free energies of the conformations as a function of accessible surface. The data from all 851 $\times$ 3 conformations are shown (a) (see the caption of Fig. 3.7). The data for (b) and (c) are from the conformations which have $E_{TOT} < 40$ kcal/mol in Run 1 and Run 2, respectively.

(b)



(c)

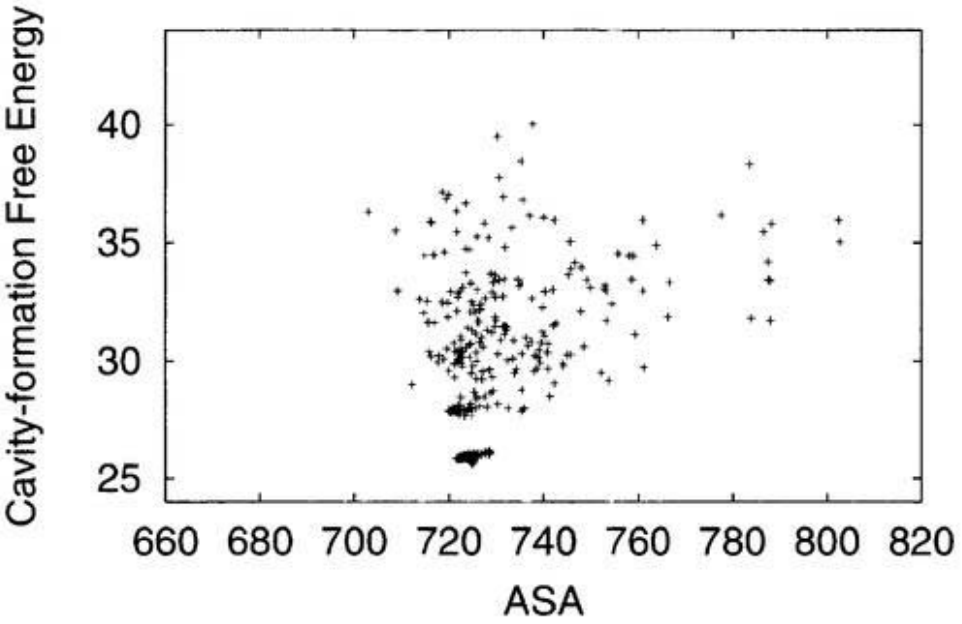


Fig. 3.8 (Continued)

Bibliography

[1] G. H. Paine and H. A. Scheraga, *Biopolymers* **26**, 1125 (1987).

[2] Z. Li and H. A. Scheraga, *Proc. Natl. Acad. Sci. USA* **84**, 661 (1987).

[3] H. Kawai, T. Kikuchi,and Y. Okamoto, *Protein Eng.* **3**, 85 (1989).

[4] B. Von Freyberg and W. Braun, *J. Comput. Chem.* **21**, 1065 (1991).

[5] Y. Okamoto, T. Kikuchi, and H. Kawai, *Chem. Lett.* **1992**, 1275.

[6] U. H. E. Hansmann and Y. Okamoto, *J. Comp. Chem.* **14**, 1333 (1993).

[7] U. H. E. Hansmann and Y. Okamoto, *Physica A* **212**, 415 (1994).

[8] H. Meirovitch, E. Meirovitch, A. G. Michel, and M. Vásquez, *J. Phys. Chem.* **98**, 6241 (1994).

[9] F. Eisenmenger and U. H. E. Hansmann, *J. Phys. Chem. B* **101**, 3304 (1997).

[10] I. P. Androulakis, C. D. Maranos, and C. A. Floudas, *J. Global. Opt.* **11**, 1 (1997).

[11] L. Carlacci, *J. Comput. Aided. Mol. Des.* **12**, 195 (1998).

[12] S. Kirkpatrick, C.D. Gelatt, Jr., and M. P. Vecchi, *Science* **220**, 671 (1983).

[13] M. Irisa, T. Takahashi, F. Hirata, and T. Yanagida, *J. Mol. Liquids.* **65**, 381 (1995).

[14] M. Irisa, T. Takahashi, K. Nagayama, and F. Hirata, *Mol. Phys.* **85**, 1227 (1995).

[15] H. Reiss, H. L. Fish, and J. L. Lebowitz, *J. Chem. Phys.* **31**, 369 (1959).

[16] R. A. Pierotti, *Chem. Rev.* **76**, 717 (1965).

[17] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. J. Teller, *J. Chem. Phys.* **21**, 1087 (1953).

[18] Y. Okamoto, M. Fukugita, T. Nakazawa, and H. Kawai, *Protein Eng.* **4**, 639 (1991).

[19] Y. Okamoto, T. Kikuchi, and H. Kawai, *Chme. Lett.* **1993**, 1275.

[20] M. Irisa, *Comput. Phys. Comm.* **98**, 317 (1996).

[21] M. Kinoshita, Y. Okamoto, and F. Hirata, unpublished.

Chapter 4

Temperature Dependence of Distributions of Conformations of Met-Enkephalin in Gas Phase

Ayori Mitsutake, Ulrich H.E.Hansmann, and Yuko Okamoto, "Temperature Dependence of Distributions of Conformations of a Small Peptide," *The Journal of Molecular Graphics and Modelling* **16**, 226-238 (1999).

4.1 Introduction

In the previous Chapter, we studied the low-energy conformations of Met-enkephalin by Monte Carlo simulated annealing [1]. Simulated annealing is powerful and easy to implement, but information only in the vicinity of the lowest-energy state can be obtained by this method. In order to study the folding of the peptide from a random coil state, the detailed, systematic analyses of the conformations as a function of temperature have to be performed. This is the purpose of the present chapter. The algorithm we adopt is multicanonical algorithm [2], which allows one to obtain various thermodynamic quantities as a function of temperature. In this work we first classify the low-energy conformations of Met-enkephalin into several groups of similar structures. We then present the distributions of conformations, hydrogen bonds, and dihedral angles as a function of temperature, which gives much information about the free energy landscape of Met-enkephalin.

4.2 Computational Details

In this work we adopted multicanonical simulation to obtain various thermodynamic quantities as a function of temperature. Only the conformational energy term (E_P) (see the subsection 2.1.1) [3, 4, 5] is estimated, because we want to compare the results by multicanonical simulation in the present work with those by simulated annealing simulation in gas phase. For Met-enkephalin (whose amino-acid sequence is Tyr-Gly-Gly-Phe-Met), the number of degrees of freedom is 19. One Monte Carlo (MC) sweep consists of updating all these 19 angles once with Metropolis evaluation [6] for each update. For our study of Met-enkephalin, we first made a preliminary canonical simulation at T_0=1000 K with 100 000 Monte Carlo sweeps. We iterated this process thirteen times to determine the optimal weighting factor $w_{mu}(E)$ by using version 1 in subsection 2.2.2. After the optimal weighting factor $w_{mu}(E)$ was determined we then made one production run with 1 000 000 Monte Carlo sweeps and obtained various thermodynamic quantities as a function of temperature by the reweighting techniques [7]. Analysis of the time series of energy showed that the present choice of multicanonical weight factor indeed realizes a random walk in

potential energy space, which keeps the simulation from getting trapped in a local minimum (see Fig. 4.1). The random walk visits the lowest-energy region ($E \approx -12$ kcal/mol) several times in 1 000 000 Monte Carlo sweeps. The visits are separated by excursions into high-energy regions, which ensures de-correlation of the configuration.

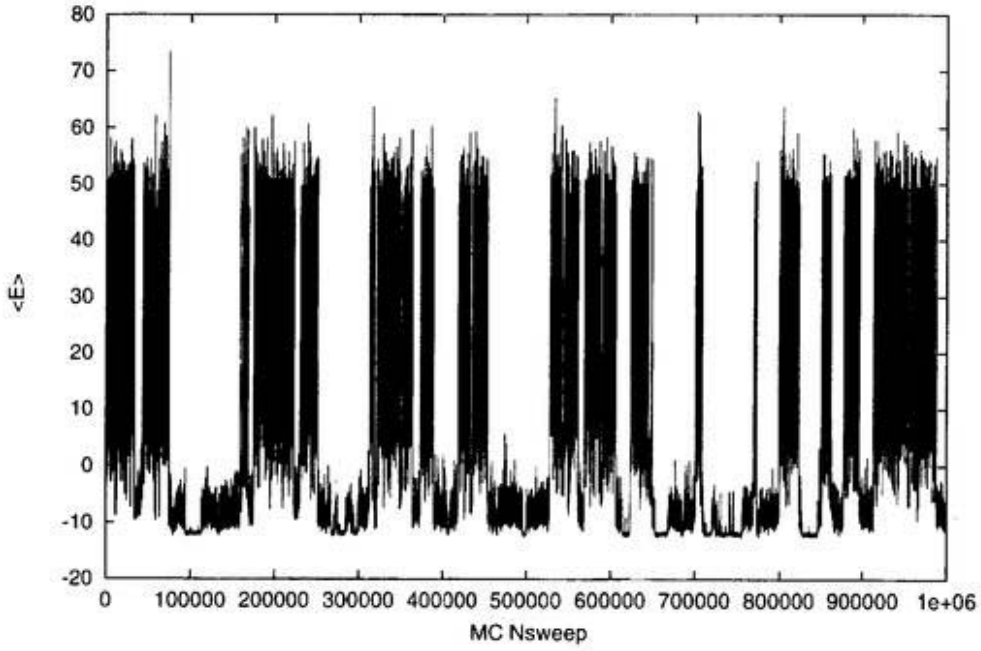


Figure 4.1: The time series of the total potential energy E_P (kcal/mol) obtained by the multicanonical production run.

The large changes in energy imply large conformational changes that occur in the course of the simulation. Because large parts of the configuration space are sampled, the use of the reweighting techniques is justified to calculate thermodynamic quantities over a wide range of temperature. As an example, we calculated the average energy as a function of temperature (see Fig. 4.2) and the results were in good agreement with those in previous multicanonical simulations.[8, 9, 10]

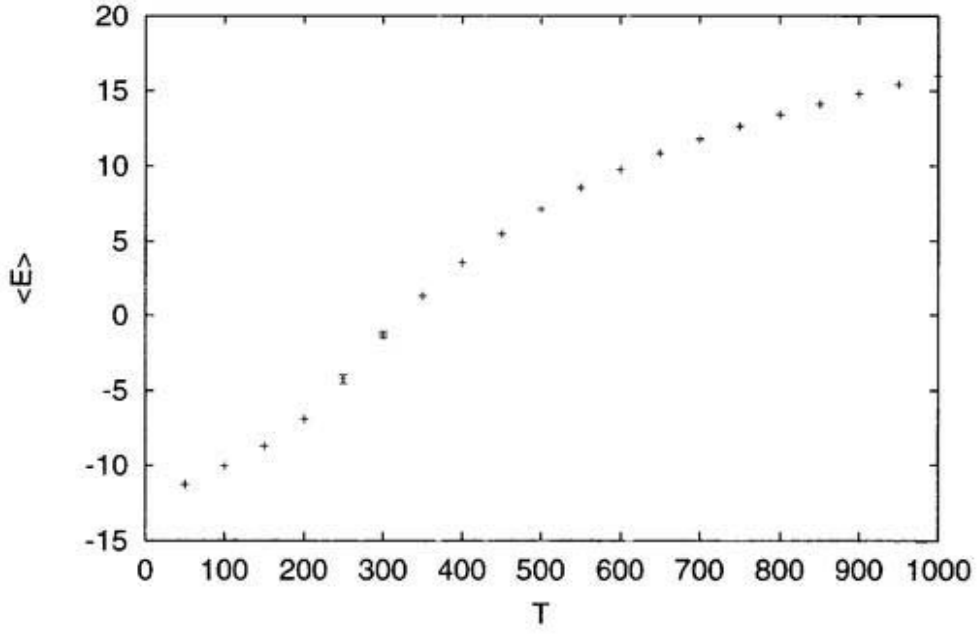


Figure 4.2: The average total potential energy as a function of temperature.

4.3 Results and Discussion

4.3.1 Distributions of Hydrogen Bonds

In order to study the characteristics of the obtained conformations in detail, we first examine the distributions of intrachain hydrogen bonds in the backbone. In this work, we use the following abbreviations for the atoms in the backbone. The amide nitrogen and carbonyl oxygen of the i-th residues are referred to as Ni and Oi, respectively. For example, O2 and N5 stand for the carbonyl oxygen of Gly-2 and the amide nitrogen of Met-5, respectively. Moreover, the intrachain hydrogen bond between Ni and Oj is denoted by Ni-Oj. Let d_{HO} be the distance between the amide hydrogen and the carbonyl oxygen, and θ_{NHO} the angle spanned by the amide nitrogen, amide hydrogen, and carbonyl oxygen (i.e., the angle between two vectors $\vec{NH}$ and $\vec{HO}$). We adopt the following criterion for hydrogen-bond formations. We consider that a hydrogen bond between the amide nitrogen (donor) and carbonyl oxygen (acceptor) is formed if

$$\begin{cases} d_L \leq d_{HO} \leq d_H \ , \\ \theta_{NHO} \leq \theta_C \ , \end{cases} \tag{4.1}$$

where we take

$$d_L = 1.5 \text{ Å}, \ d_H = 2.5 \text{ Å}, \text{ and } \theta_C = 60^\circ \ . \tag{4.2}$$

Since adjacent residues do not form hydrogen bonds, there are 12 possible hydrogen-bond patterns (namely, 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). Using the reweighting techniques, we have calculated the probability distributions of these hydrogen-bond formations as a function of temperature. The results are shown in Fig. 4.3.

It is found in Fig. 4.3 that the distributions of the hydrogen bonds O2-N5 and N2-O5 are very high at low temperatures and monotonically decreasing as the temperature increases. More than 90 % of the conformations have these two hydrogen bonds below 150 K. As the temperature is raised, four more hydrogen bonds (N1-O4, O1-N4, O1-N3, and O2-N4) appear. The temperatures of the peak of the distributions are about 300, 300, 350, and 350 K, for N1-O4, O1-N4, O1-N3, and O2-N4 respectively. As the temperature is further raised, other hydrogen bonds will also appear, but their contributions are rather small.

4.3.2 Classification of Conformations

As was done in earlier work [11, 12, 10], the patterns of intrachain hydrogen bonds can naturally be used for classification of conformations in a small peptide such as Met-enkephalin. In particular, we try to classify the structures by the patterns of hydrogen bonds that connect a pair of residues. For this purpose we introduce the following notations: We say that the hydrogen bond R_i-R_j between a pair of residues R_i and R_j ($j > i+1$) is formed when at least one of the hydrogen bonds Ni-Oj and Oi-Nj are formed. For instance, R_1-R_3 can mean that N1-O3 or O1-N3 or both N1-O3 and O1-N3 are formed. There are six such hydrogen bonds (namely, R_1-R_3, R_1-R_4, R_1-R_5, R_2-R_4, R_2-R_5, and R_3-R_5). We can therefore classify the obtained conformations into six groups of similar structures, which we respectively refer to as C13, C14, C15, C24, C25, and C35. Schematic drawings of these six groups are shown in Fig. 4.4. A conformation can in principle have more than one hydrogen-bond pattern R_i-R_j, and then we have an ambiguity as to which group the conformation belongs to. For example, a conformation can have

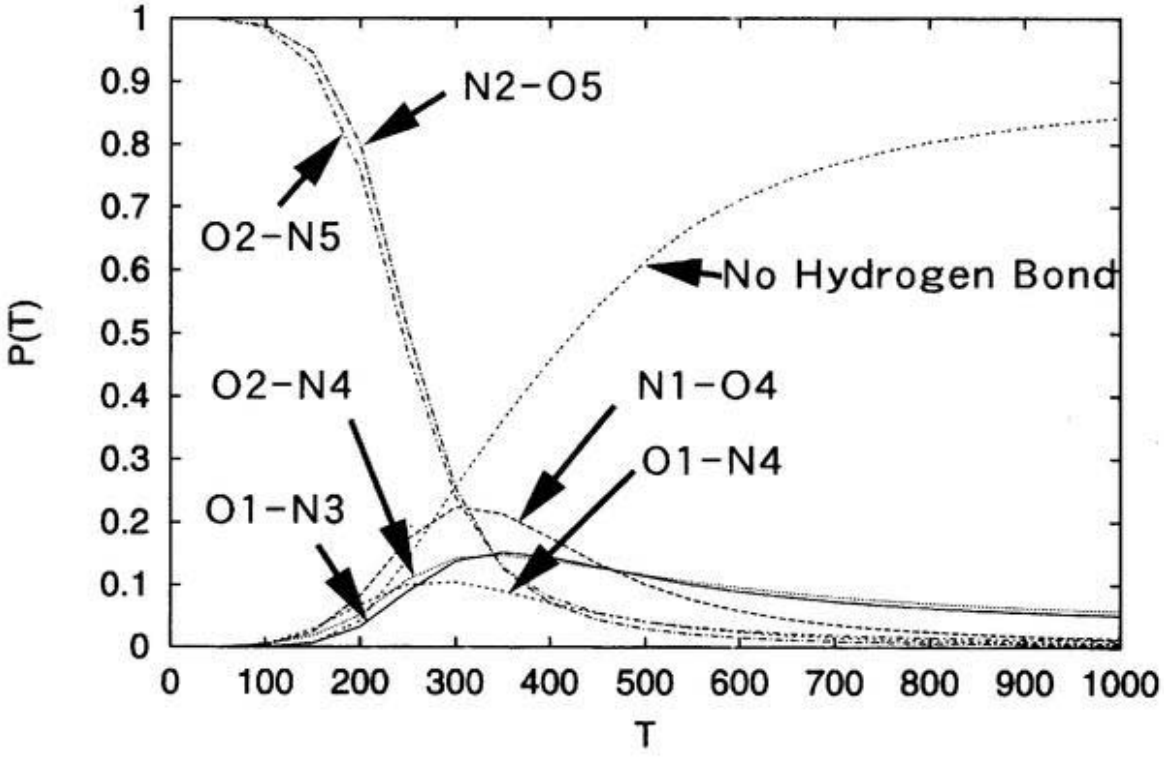


Figure 4.3: The distributions of hydrogen bonds as a function of temperature.

two hydrogen bonds, R_2-R_4 and R_2-R_5, and the conformation can belong to either group C24 or group C25. In these cases we use the most external hydrogen bond for the classification because external hydrogen-bonds will constrain the structure more than internal ones. For instance, if a conformation has two types of hydrogen bonds R_1-R_3 and R_1-R_5, we consider that the conformations belongs to groups C15, because the hydrogen bond R_1-R_3 does not restrict the structure of the C-terminal residue Met-5. The exception is the case when a conformation has hydrogen bonds R_1-R_4 and R_1-R_5. By examining such conformations obtained by the present simulation in detail, we found that their backbone structures have more similarity to group C14 than to group C15. We thus consider that the conformation belongs to group C14 instead of group C15.

As discussed in the preceding subsection, the conformations were classified into four groups (A, B, C, and D) by using the root-mean-square distances in the previous work [13]. In the present criterion of classification, the groups A, B, C, and D actually correspond

to the groups C25, C14, C15, and C24, respectively. Similarly, groups C25 and C14 correspond to structures (2),(b) and (1),(a), respectively, in Ref. [10].

After the definition of groups of similar structures is given, we now study the distributions of conformations in these groups as a function of temperature T. The results are shown in Fig. 4.5. As can be seen in the Figure, group C25 is dominant at low temperatures. Conformations of group C14 start to appear from $T \approx 100$ K. At $T \approx 300$ K, the distributions of these two groups, C25 and C14, balance (≈ 25 % each) and constitute the main groups. Above $T \approx 300$ K, the contributions of other groups become non-negligible (those of group C24 and group C13 are about 10 % and 8 %, respectively, at $T = 400$ K). Note that the distribution of conformations that do not belong to any of the six groups monotonically increases as the temperature is raised. This is because random-coil conformations without any intrachain hydrogen bonds are favored at high temperatures. These results are in agreement with Refs. [9, 10].

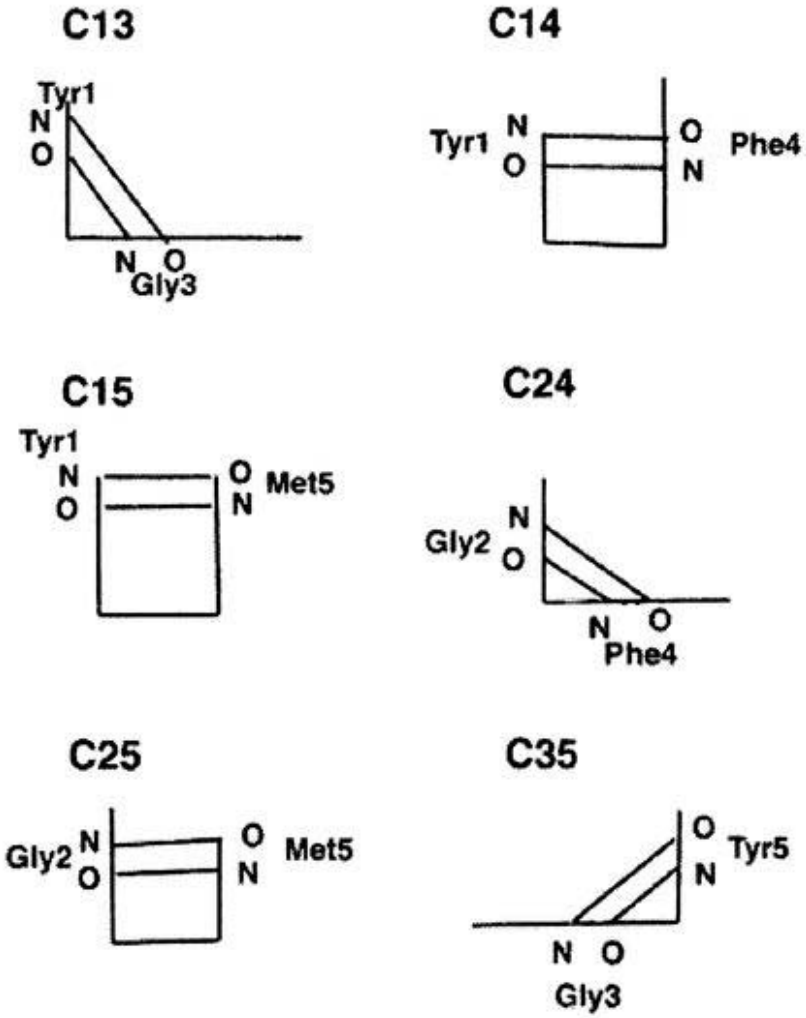


Figure 4.4: Schematic drawings of six groups of similar structures of Met-enkephalin.

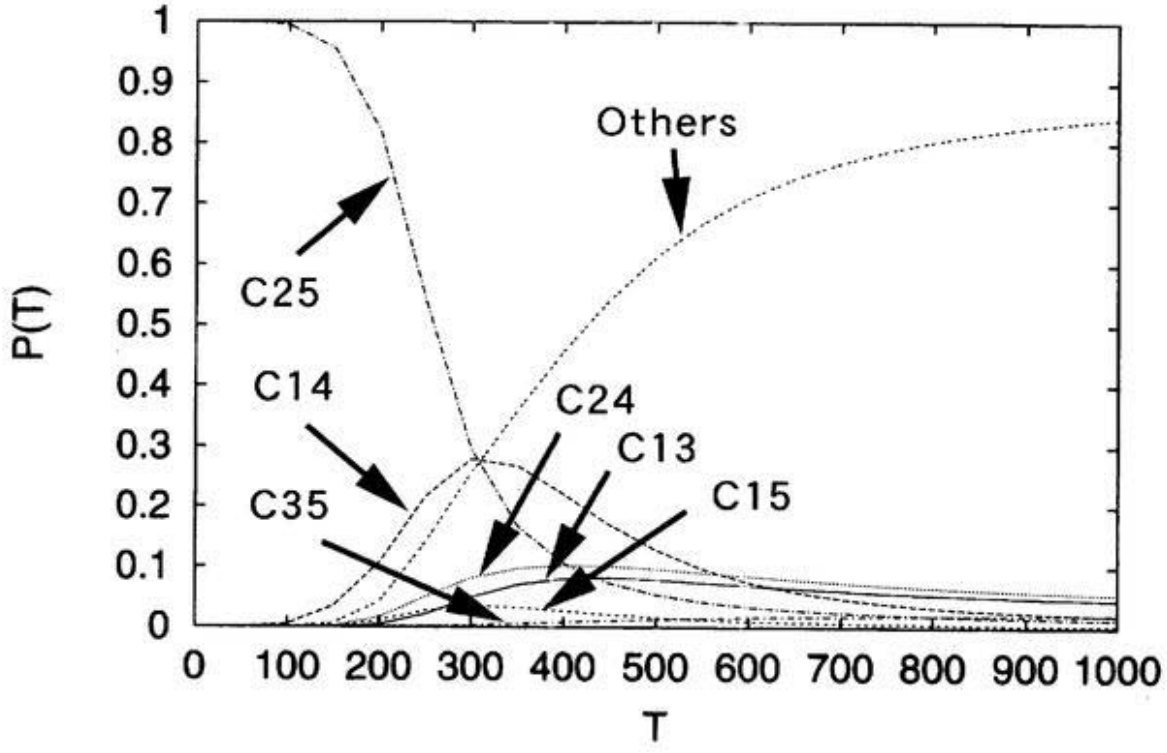


Figure 4.5: The distributions of each group of similar structures as a function of temperature.

4.3.3 Details of the Six Classified Structures

After the distributions of the six groups are established, we now study the characteristics of each group in detail. In Fig. 4.6 we show the lowest-energy conformations in each group. Conformations were obtained by minimizing (with Newton-Raphson method) the lowest-energy conformations found in each group during the multicanonical simulation. We remark that the minimization did not alter the structure much at all, but energy values were lowered from -12.2, -10.1, -8.3, -6.0, -7.0, and -1.7 kcal/mol to -12.2, -11.1, -9.8, -9.1, -8.8, and -5.0 kcal/mol for groups C25, C14, C24, C13, C15, and C35, respectively. The four of these structures (C25, C14, C24, and C15) in Fig. 4.6 are essentially identical with the lowest-energy conformations of the four groups found previously in Ref. [13] including the side-chain structures, although the previous classification used root-mean-square distances instead of hydrogen-bond patterns.

As is shown in Fig. 4.6(a), the conformation of group C25 has two hydrogen bonds, N2-O5 and O2-N5, and forms a type II' β-turn. The energy of the conformation is -12.2 kcal/mol, and this conformation corresponds to the global-minimum-energy state of Met-enkephalin in gas phase. The conformation of group C14 (Fig. 4.6(b)) has two hydrogen bonds, N1-O4 and O1-N4, and forms a type II β-turn. The energy is -11.1 kcal/mol, and this conformation corresponds to the second-lowest-energy state. The groups C25 and C14 are the dominant groups in the energy landscape of Met-enkephalin as shown above (see Fig. 4.5).

The conformation of group C24 (Fig. 4.6(c)) has only one hydrogen bond, O2-N4, (because the distance between N2 and O4 is larger than 2.5 Å) and forms a γ-turn. The energy is -9.9 kcal/mol, and this conformation corresponds to the third-lowest-energy state. We remark that the structures of group C24 can be obtained from that of group C25 by small rearrangements of the intrachain hydrogen bonds. In group C25 of Fig. 4.6(a), O2 and N4 are already close to each other. By cutting the two hydrogen bonds, N2-O5 and O2-N5, we can easily lead the conformation to that of group C24. Note that the OH of Tyr-1 side chain is hydrogen-bonded to the carbonyl oxygen of Gly-3 in both groups C25 and C24.

The conformation of group C13 (Fig. 4.6(d)) has only one hydrogen bond, O1-N3, and also forms a γ-turn. The energy is -9.1 kcal/mol and this conformation corresponds to the fourth-lowest-energy state. The structure of group C13 can be obtained from that of group C14 by small rearrangements of the intrachain hydrogen bonds. In group C14 of Fig. 4.6(b), O1 and N3 are already close to each other (see Fig. 4.6(b) also). By cutting the two hydrogen bonds, N1-O4 and O1-N4, we can easily lead the conformations to that of group C13.

The conformation of group C15 (Fig. 4.6(e)) has only one hydrogen bond, O1-N5. the energy is -8.8 kcal/mol and this conformation corresponds to the fifth-lowest-energy state. As can be seen in Fig. 4.5, the probability distribution for this group is rather low compared to the four previous groups (C25, C14, C24, and C13).

The conformation of group C35 (Fig. 4.6(f)) has only one hydrogen bond, O3-N5. The backbone of this conformation is extended and the energy is rather high (-5.0 kcal/mol).

As can been seen in Fig. 4.5, the probability distribution for this group is also low compared to the first four groups (C25, C14, C24, and C13).

Finally, in Table 4.1 we list the root-mean-square distances between pairs of the six conformations in Fig. 4.6. Only the atoms in the backbone are taken into account. The entries are all more than or equal to 3.0 Å, and this implies that these structures are indeed quite different.

Table 4.1: Root-mean-square distances (Å) of the coordinates of the backbone atoms among the lowest-energy conformations[a] in each of the six groups (C13, C14, C15, C24, C25, and C35).

	C13	C14	C15	C24	C25	C35
C13	0.0	3.3	3.2	3.1	5.2	3.1
C14	3.3	0.0	4.3	3.0	4.1	4.0
C15	3.2	4.3	0.0	3.4	4.7	3.0
C24	3.1	3.0	3.4	0.0	3.9	3.1
C25	5.2	4.1	4.7	3.9	0.0	5.1
C35	3.1	4.0	3.0	3.1	5.1	0.0

[a] These six conformations are shown in Fig. 4.6.

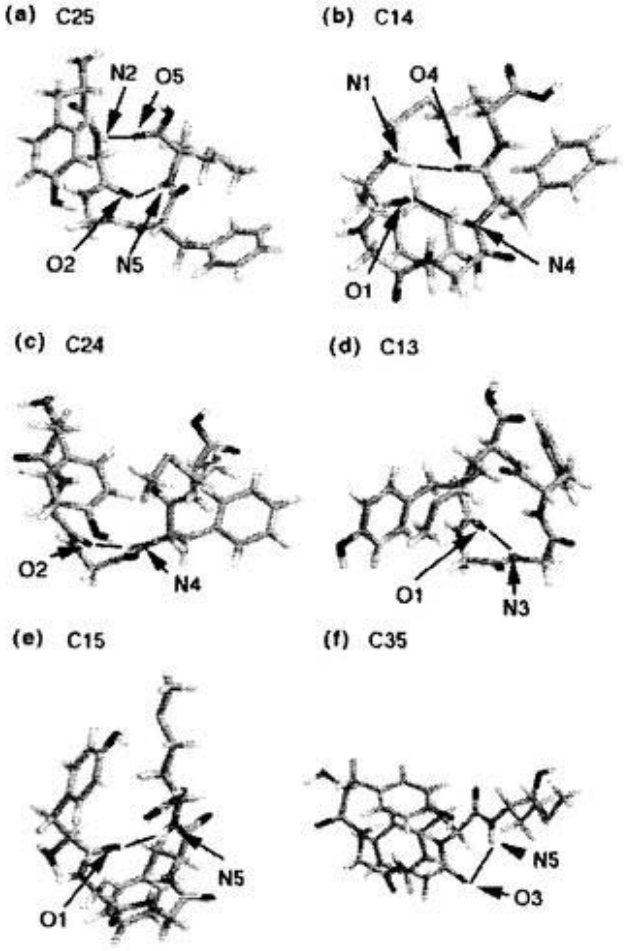


Figure 4.6: The lowest-energy conformations in each group obtained by the present simulation. These conformations were obtained by minimizing (with Newton-Raphson method) the lowest-energy conformations found during the multicanonical simulation. The lowest-energy conformations correspond to groups C25(a), C14(b), C24(c), C13(d), C15(e), and C35(f)

4.3.4 Hydrogen Bond formations in Each Group of Similar Structures

In Fig. 4.3 we saw the distributions of all possible intrachain hydrogen bonds as a function of temperature. We now examine the possible hydrogen bonds as a function of temperature for each group separately in order to study how the conformations in each group are disordered as the temperature is raised. The results are shown in Fig. 4.7. We only show

results for groups C25 and C14, because these groups are dominant. Here, we used the following weaker conditions for hydrogen-bond formations in Eq. (4.1) in order to judge which group each conformation belongs to:

$$d_L = 1.5 \text{ Å}, \; d_H = 3.0 \text{ Å}, \text{ and } \theta_C = 90^\circ \; , \tag{4.3}$$

because we want to include conformations that do not have hydrogen bonds in a strict sense but yet have similar structures (and the actual distributions of hydrogen bonds were calculated by the definition with Eq. (4.2)).

For group C25 (Fig. 4.7(a)), the distributions of the two characteristic hydrogen bonds, N2-O5 and O2-N5, are equally high at low temperatures. This implies that the lowest-energy conformation in Fig. 4.6(a) is quite stable. We remark that at very high temperatures hydrogen bond O2-N5 is stronger than N2-O5, suggesting that in group C25 the external hydrogen bond (N2-O5) is easier to break than the internal one (O2-N5). The third hydrogen bond, O2-N4, also appears at higher temperatures when the conformation is slightly deformed by thermal fluctuations, opening a possibility of transition of structures from group C25 into group C24 as discussed above.

For group C14 (Fig. 4.7(b)), the distributions of the two characteristic hydrogen bonds, N1-O4 and O1-N4, are high but not equal at low temperatures. At very low temperatures about 80 % of the conformations have the two hydrogen bonds and the remaining 20 % have only one hydrogen bond, O1-N4. We remark that at higher temperatures hydrogen bond N1-O4 is stronger than O1-N4, implying that in group C14 the internal hydrogen bond (O1-N4) is easier to break than the external one (N1-O4). The third hydrogen bond, O1-N3, also appears at high temperatures, suggesting that a transition of structures from group C14 to group C13 is possible when the conformation is slightly deformed by thermal fluctuations as discussed above.

For groups C24 and C13 (data not shown), conformations have only one characteristic hydrogen bond (respectively N2-O4 and O1-N3) at low temperatures and the other hydrogen bond (respectively O2-N4 and N1-O3) does not exist, because they both form a γ-turn.

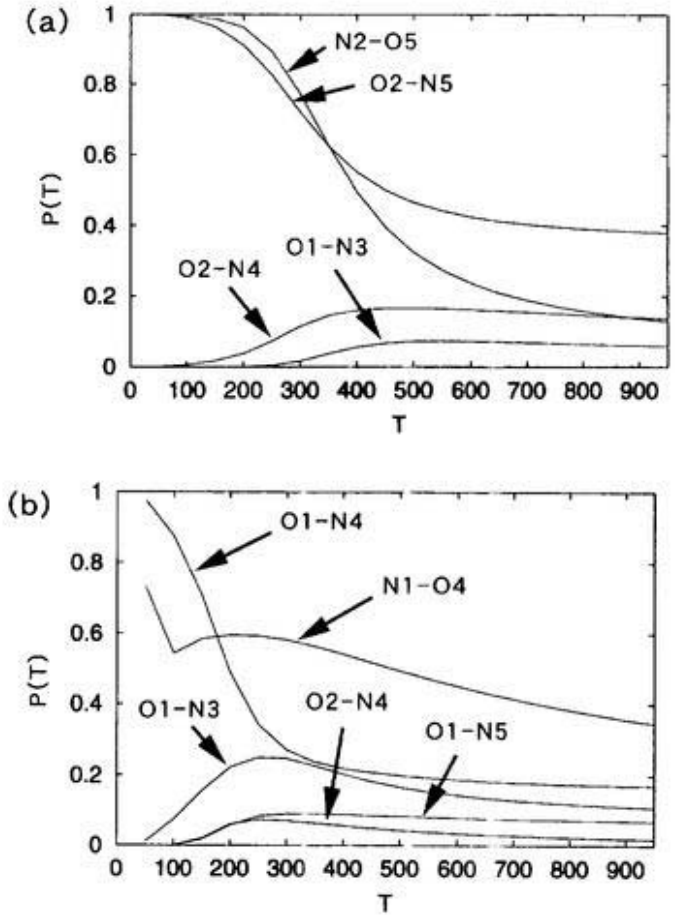


Figure 4.7: The distributions of 12 possible hydrogen-bond patterns in groups C25 and C14 as a function of temperature.

4.3.5 Distributions of Backbone Dihedral Angles

Our analyses so far imply that there are two very stable structures (groups C25 and C14) at low temperatures. They respectively form type II' β-turn and type II β-turn. There are two more stable structures (groups C24 and C13) which form γ-turns. The remaining two groups (C15 and C35) turned out to be less stable. These results imply that there are at least four well-defined valleys in the energy landscape of Met-enkephalin in gas phase. We now study the distributions of backbone dihedral angles. For each group we separately make Ramachandran plots of conformations obtained in the course of the multicanonical simulations. Here, we again use the weaker definition (Eq. (4.3))

for hydrogen bonds in Eq. (4.1) to judge which group each conformation belongs to. Two sets of Ramachandran plots are made for each residue. The first set is taken from conformations with energy E in the range $E_{min} \leq E < E_{min} + 3$ (kcal/mol) and the other set is from the range $E_{min} + 3 \leq E < E_{min} + 6$ (kcal/mol), where E_{min} stands for the lowest-energy in each group obtained during the multicanonical simulation. (Namely, $E_{min} = -12.2,\ -10.1,\ -8.3,\ -6.0,\ -7.0$, and -1.7 kcal/mol for groups C25, C14, C24, C13, C15, and C35, respectively.) We consider that the group in question is a well-defined valley in the energy landscape if the second set of Ramachandran plots are distributed more widely than the first set and contain the first set as a subset.

The results are shown in Fig. 4.8. We only show data for groups C25, C14, C24, and C13. For groups C25 (Fig. 4.8(a)) and C14 (Fig. 4.8(b)), the numbers of samples are large, and the above criterion is certainly met. These groups are thus well-defined valleys in the energy landscape. This can also be checked directly by minimizing the configurations. They all converge into the same local-minimum state. For group C25 the distributions are well localized except for ϕ of Tyr-1 (ϕ_1). This implies that the structure is very stable. For ϕ_1 the angle is free to vary because it is located at the end of the peptide and no hydrogen bond stabilizes the structure there. On the other hand, for group C14 the C-terminal dihedral angles (ϕ_5, ψ_5) are free to move because the hydrogen bond R_1-R_4 does not constrain the Met-5 structure.

For group C24 (Fig. 4.8(c)) and group C13 (Fig. 4.8(d)), the numbers of sample dots are not as large as those for group C25 and group C14, but the criterion for valley in the energy landscape is again satisfied. These two groups can thus be said to correspond to local-minimum states. We remark that the distributions for group C24 is rather similar to those for group C25 except for those of Phe-4 and Met-5. It is therefore easy for these two structures to interchange with each other as discussed above. Note also that the structure of Met-5 is flexible for C24 because no hydrogen bond stabilizes it. The distributions for C13 is likewise similar to those for C14 except for those of Gly-3.

For C15 (Fig. 4.8(e)) and C35 (Fig. 4.8(f)), the numbers of samples are rather small and we cannot conclusively claim that these groups correspond to valleys in the energy landscape. If they do, they are shallower than those of the previous four groups.

Using the reweighting techniques, we can calculate the distributions of the backbone dihedral angles of Met-enkephalin as a function of temperature. The results at 100 K, 300 K, and 1000 K are shown in Fig. 4.9. At $T = 100$ K (Fig. 4.9(a)) the distributions are well-localized, and at this temperature one conformation (C25) is dominant. At $T = 300$ K (Fig. 4.9(b)) the distributions are still localized but we see a trace of the second conformations (group C14) appearing. Note that these results are in complete agreement with our analyses of the energy landscape of Met-enkephalin in Refs. [14, 15], where we used the overlap functions to characterize the energy landscape. There, the backbone dihedral angles themselves were taken as the order parameter and again we observed the change in the free energy landscape as a function of temperature indicating the transition between a unique ground state and an ensemble of other well-defined structures. Finally at $T = 1000$ K (Fig. 4.9(c)) the distributions are widely spread, implying the large thermal fluctuations. These results correspond to the observed transition between an ensemble of well defined compact states and extended coil structures observed in Refs. [14, 15].

Figure 4.8: Ramachandran plots in each residues of conformations obtained in the course of the multicanonical simulation. We show the results for C25(a), C14(b), C24(c), C13(d), C15(e), and C35(f). Two sets of Ramachandran plots are made for each residue. The first set is taken from conformations with energy E in the range $E_{min} \leq E < E_{min} + 3$ (kcal/mol) (left-hand side) and the other set is from the range $E_{min} + 3 \leq E < E_{min} + 6$ (kcal/mol) (right-hand side), where E_{min} stands for the lowest-energy in each group obtained during the simulation. (Namely, $E_{min} = -12.2, -10.1, -8.3, -6.0, -7.0$, and -1.7 kcal/mol for groups C25, C14, C24, C13, C15, and C35, respectively.). Abscissa is the dihedral angle ϕ and ordinate is the dihedral angle ψ.

(a)

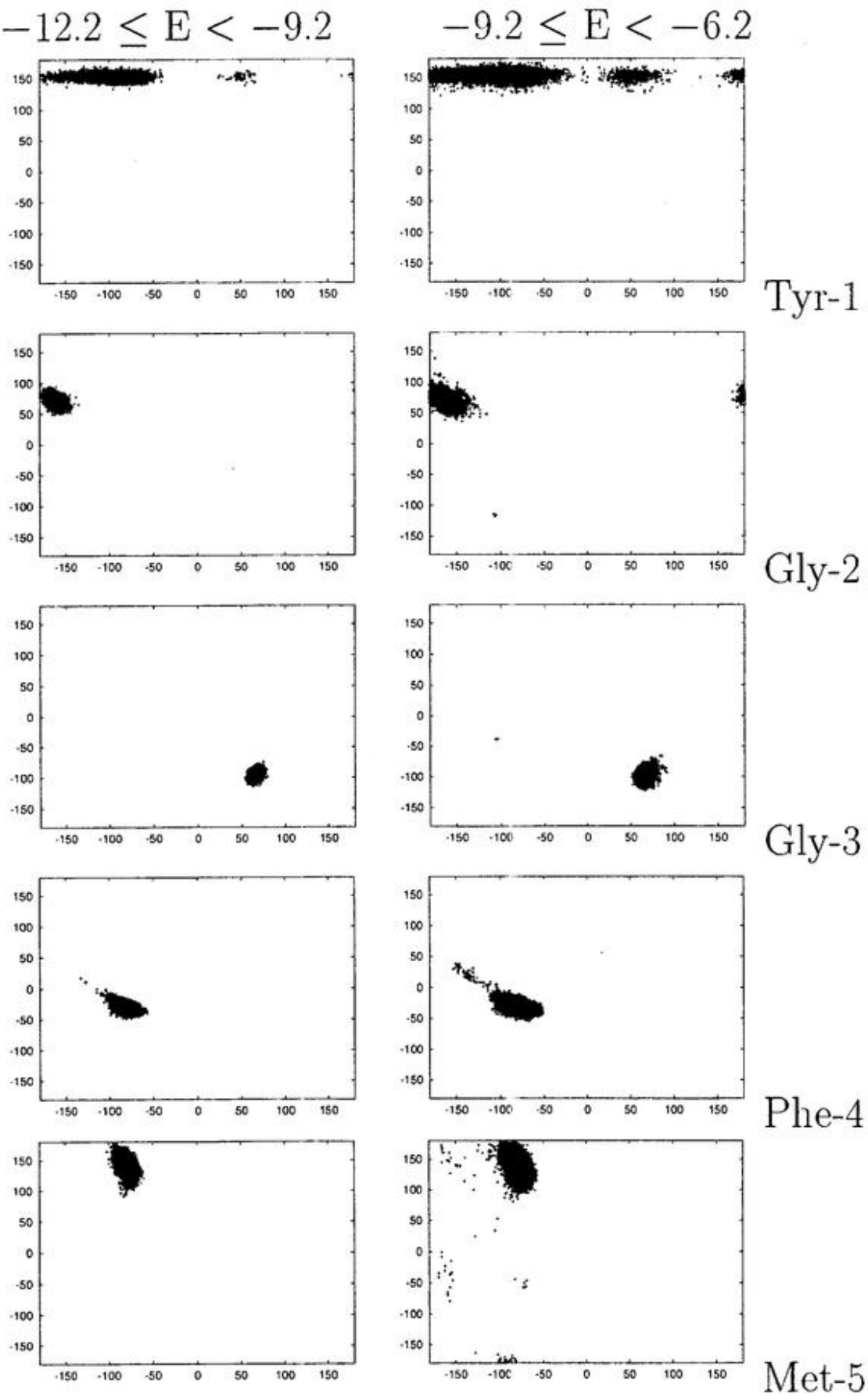


Fig. 4.8 (Continued)

(b)

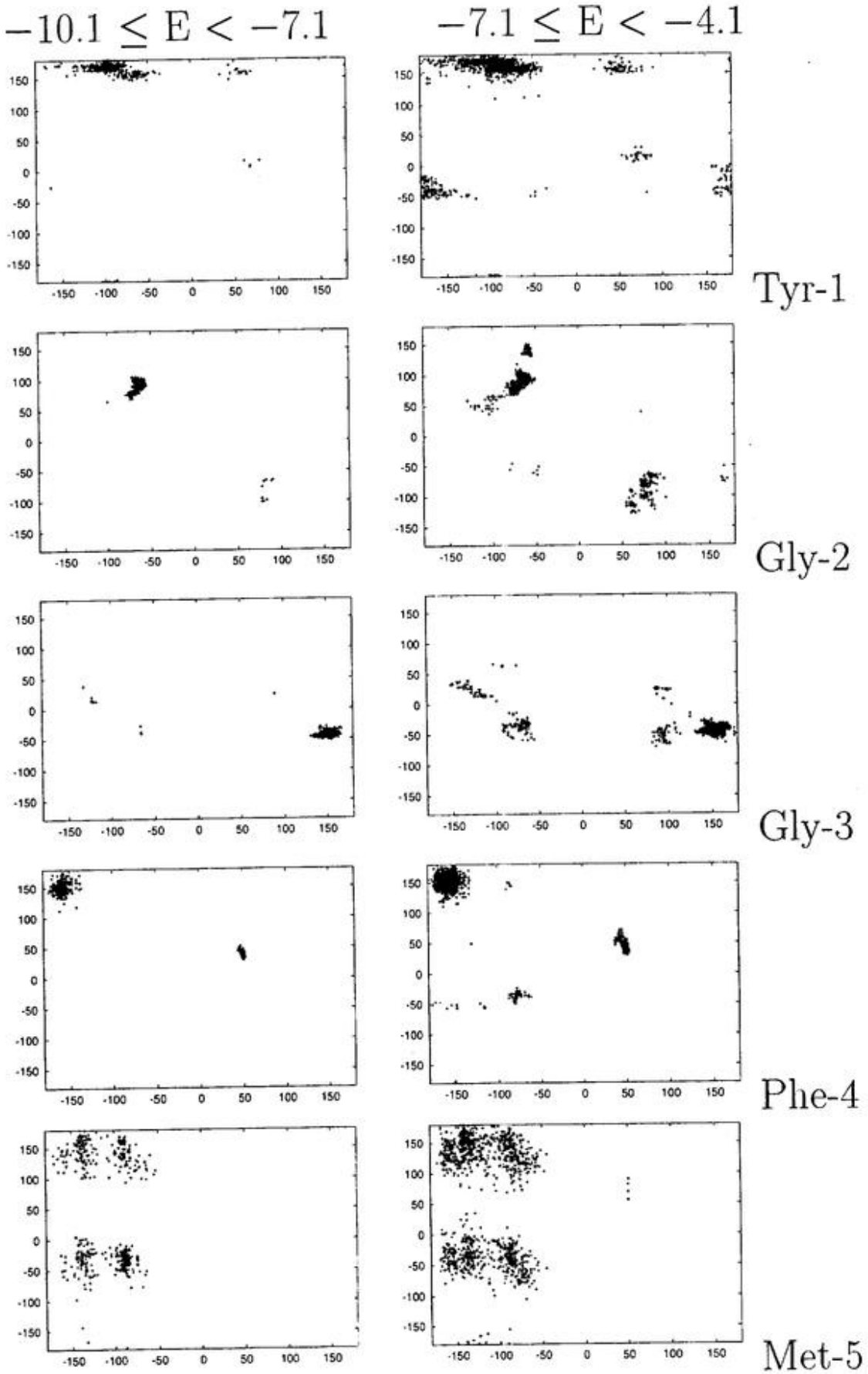


Fig. 4.8 (Continued)

(c)

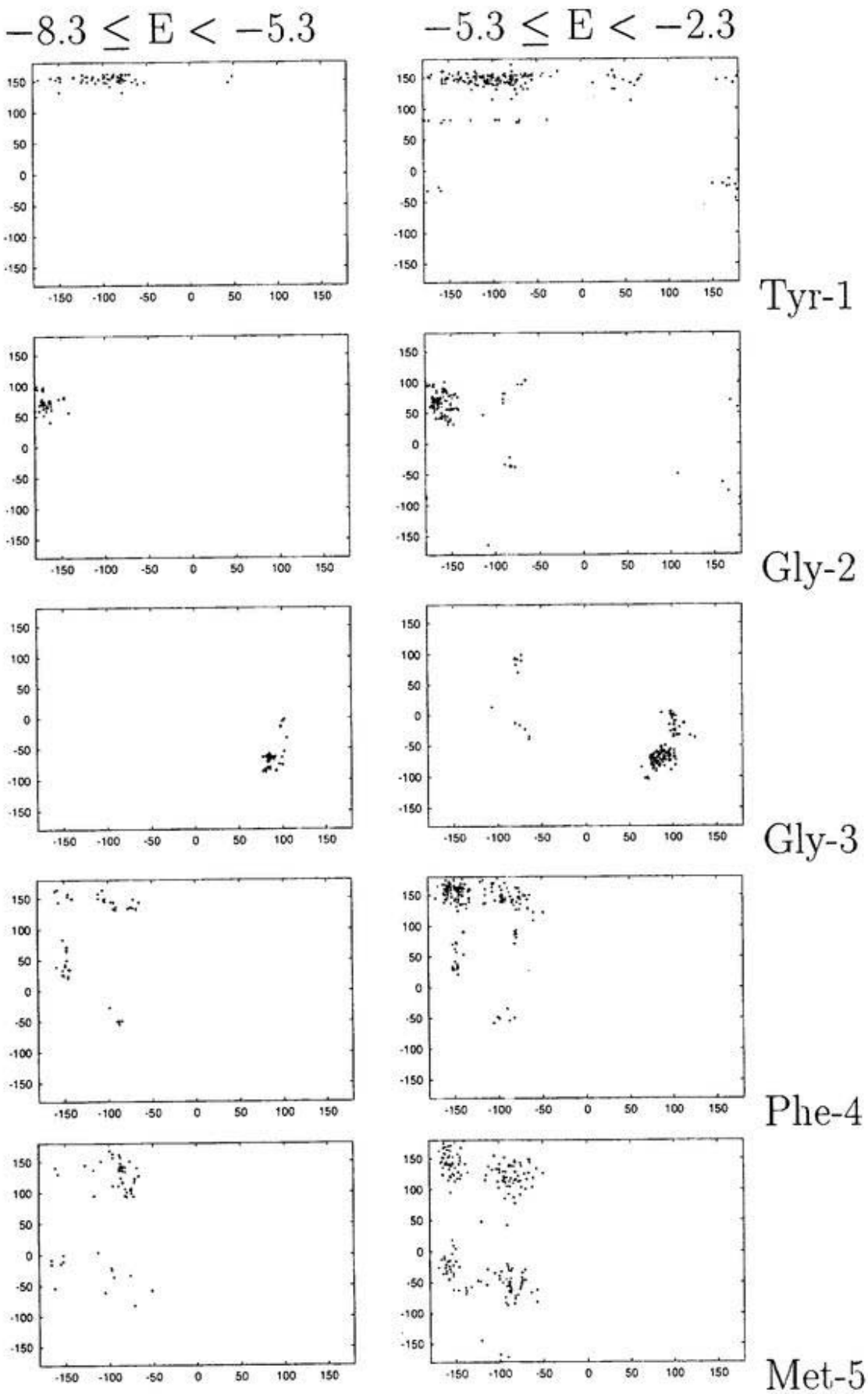


Fig. 4.8 (Continued)

(d)

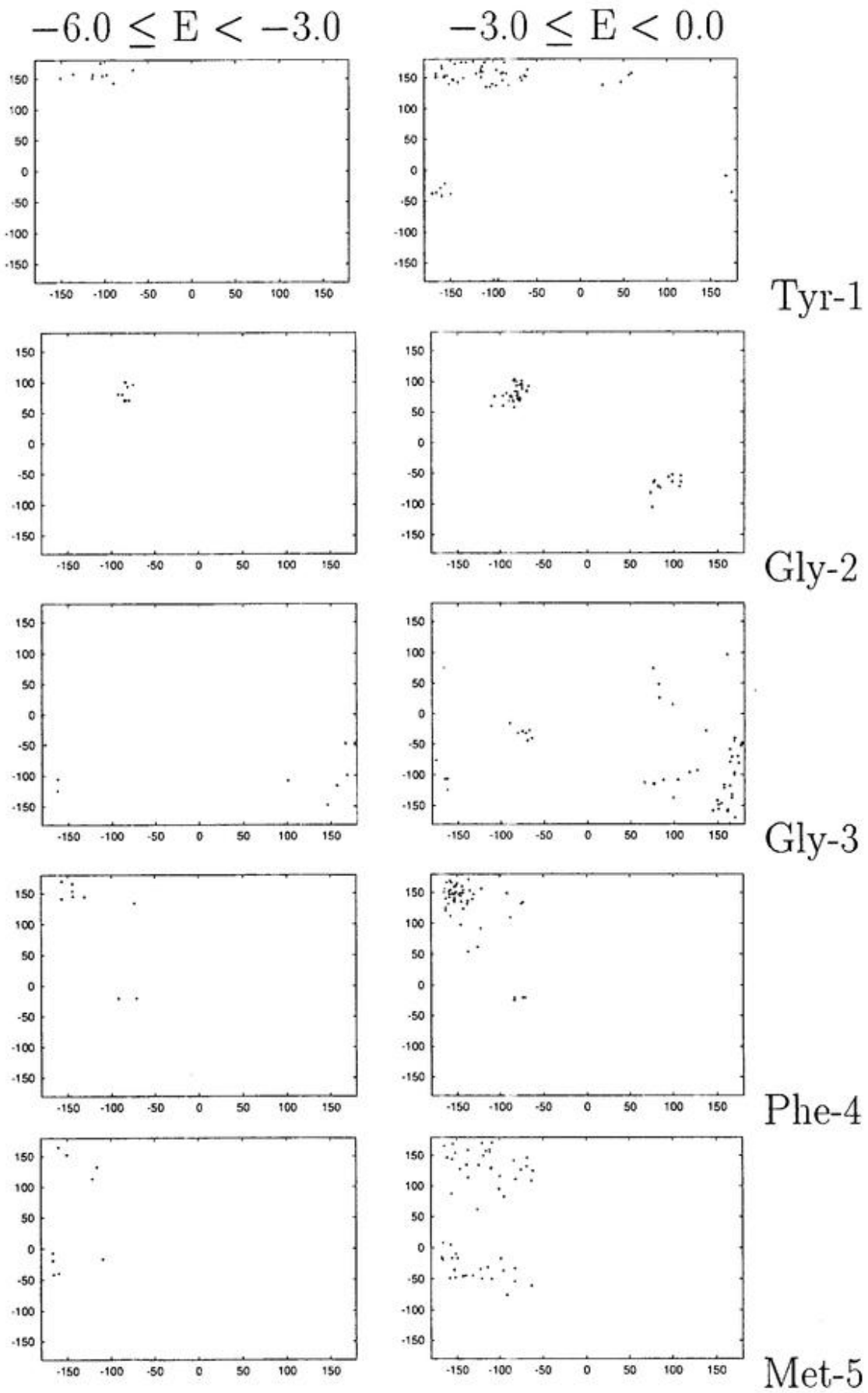


Fig. 4.8 (Continued)

(e)

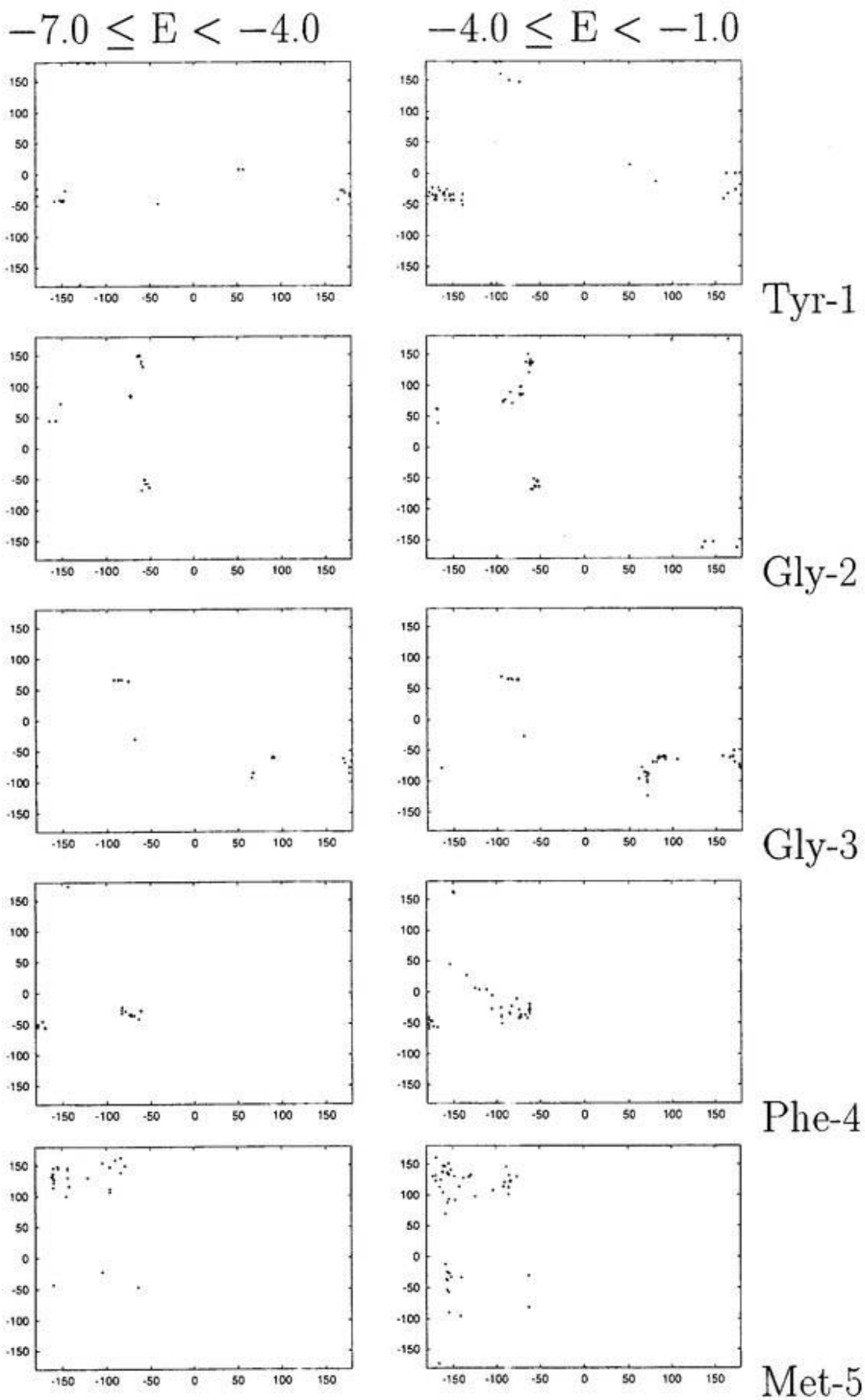


Fig. 4.8 (Continued)

(f)

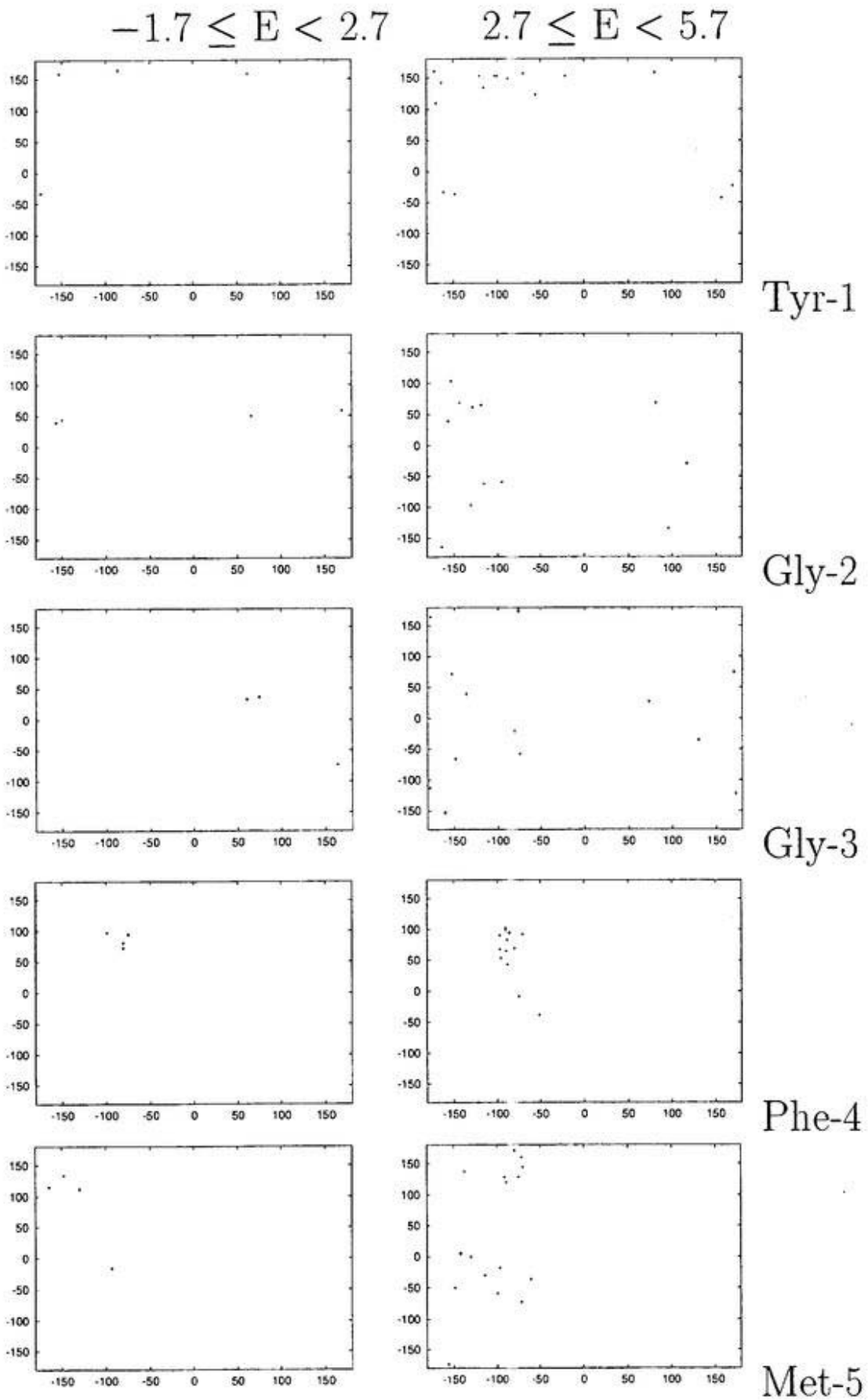
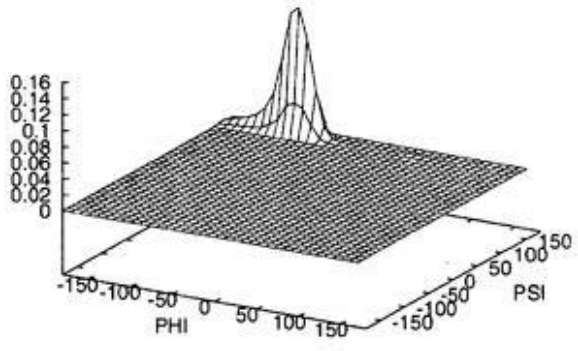


Fig. 4.8 (Continued)

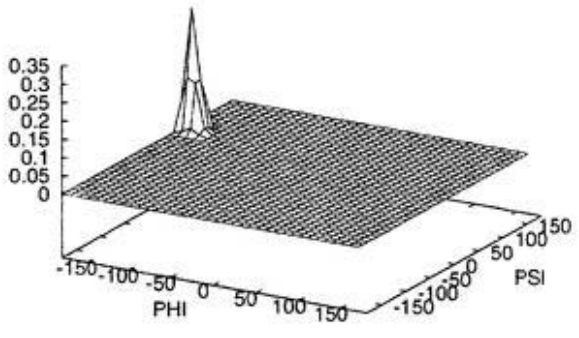
Figure 4.9: Distributions of the backbone dihedral angles of Met-enkephalin as a function of temperature. The results at $T = 100$ K(a), 300 K(b), and 1000 K(c) are shown.

(a) T=100 K

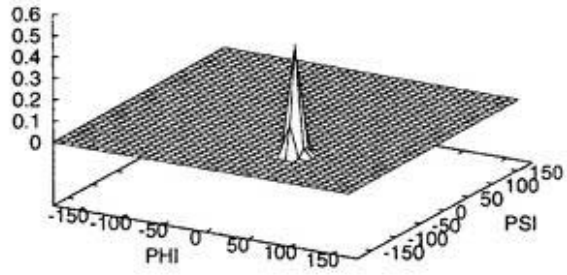
Tyr-1



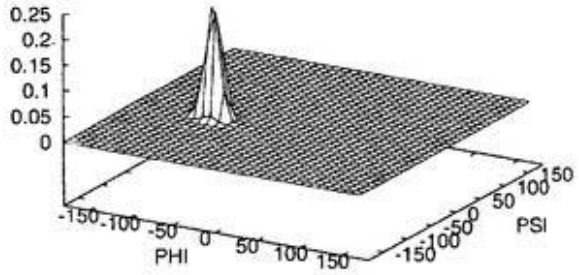
Gly-2



Gly-3



Phe-4



Met-5

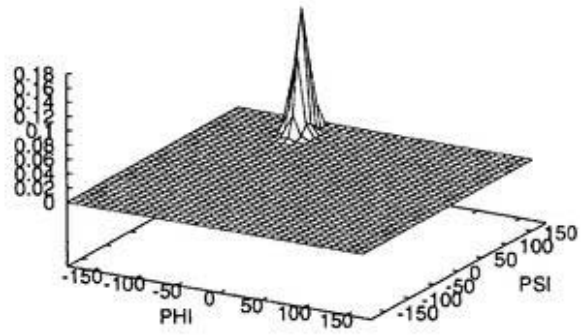
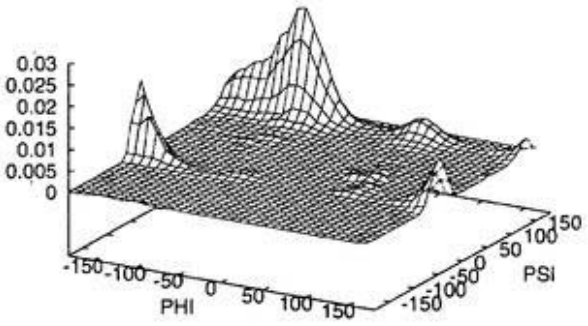


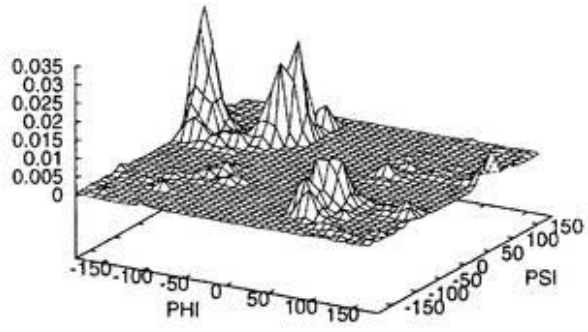
Fig. 4.9 (Continued)

(b) T=300 K

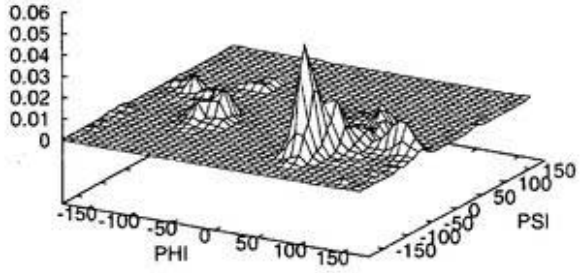
Tyr-1



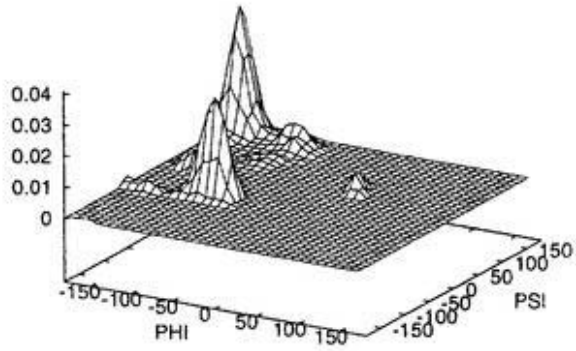
Gly-2



Gly-3



Phe-4



Met-5

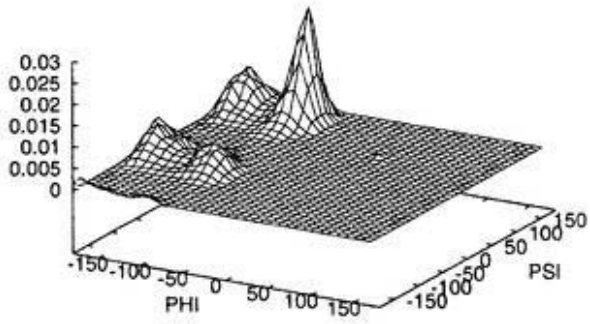
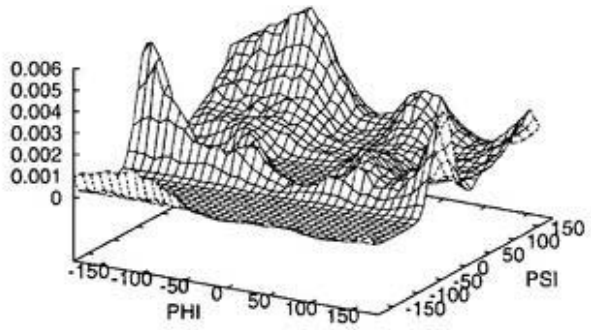


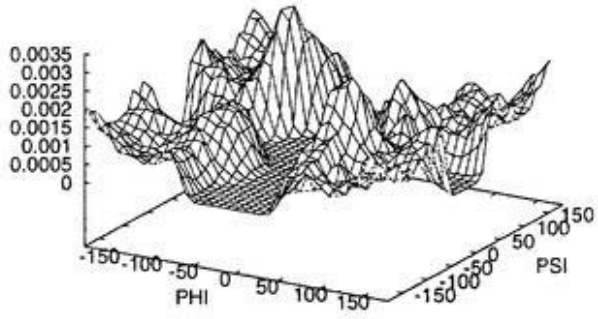
Fig. 4.9 (Continued)

(c) T=1000 K

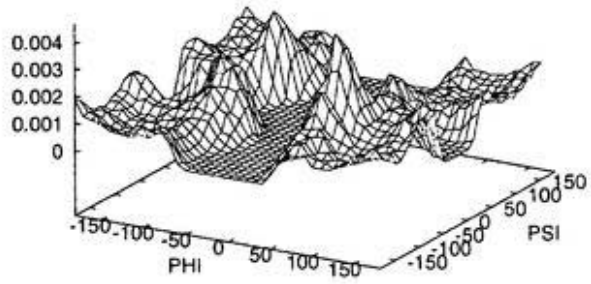
Tyr-1



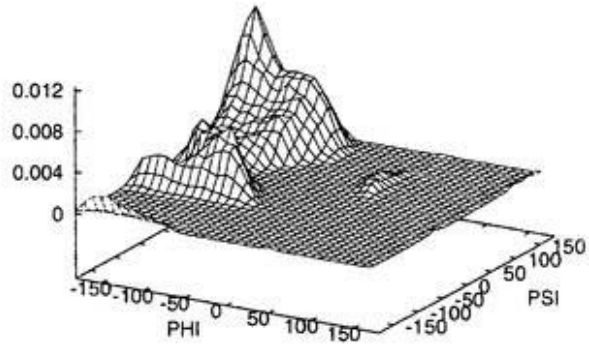
Gly-2



Gly-3



Phe-4



Met-5

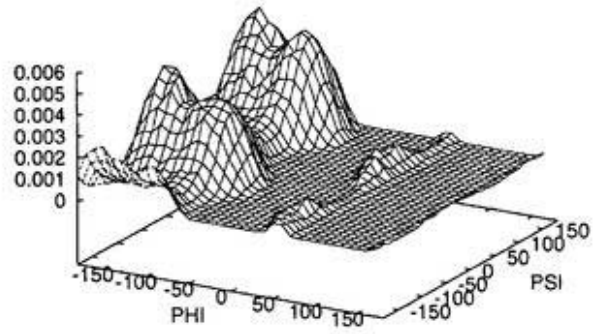


Fig. 4.9 (Continued)

Bibliography

[1] S. Kirkpatrick, C. D. Gelatt, Jr., and M. P. Vecchi, *Science* **220**, 671 (1983).

[2] B. A. Berg and T. Neuhaus, *Phys. Lett.* **B267**, 249 (1991).

[3] a) F. A. Momany, R. F. McGuire, A. W. Burgess, and H. A. Scheraga, *J. Phys. Chem.* **79**, 2361 (1975).

b) G. Némethy, M. S. Pottle, and H. A. Scheraga, *J. Phys. Chem.* **87**, 1883 (1983).

c) M. J. Sippl, G. Némethy, and H. A. Scheraga, *J. Phys. Chem.* **88**, 6231 (1984).

[4] H. Kawai, Y. Okamoto, M. Fukugita, T. Nakazawa, and T. Kikuchi, *Chem. Lett.* **1991**, 213.

[5] Y. Okamoto, M. Fukugita, T. Nakazawa, and H. Kawai, *Protein Eng.* **4**, 636 (1991).

[6] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. J. Teller, *J. Chem. Phys.* **21**, 1087 (1953).

[7] A. M. Ferrenberg and R. H. Swendsen, *Phys. Rev. Lett.* **61**, 2635 (1988).

[8] U. H. E. Hansmann and Y. Okamoto, *J. Comp. Chem.* **14**, 1333 (1993).

[9] U. H. E. Hansmann and Y. Okamoto, *Physica A* **212**, 415 (1994).

[10] F. Eisenmenger and U. H. E. Hansmann, *J. Phys. Chem. B* **101**, 3304 (1997).

[11] G. H. Paine and H. A. Scheraga, *Biopolymers* **26**, 1125 (1987).

[12] B. Von Freyberg and W. Braun, *J. Comput. Chem.* **21**, 1065 (1991).

[13] Y. Okamoto, T. Kikuchi, and H. Kawai, *Chem. Lett.* **1992**, 1275.

[14] U. H. E. Hansmann, M. Masuya, and Y. Okamoto, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 10652 (1997).

[15] U. H. E. Hansmann, Y. Okamoto, and J. Onuchic, *PROTEINS: Struct. Funct. Genet.* **34**, 472 (1999).

Chapter 5

α-Helix Propensities of Homo-Oligomers in Aqueous Solution

Ayori Mitsutake and Yuko Okamoto, "α-Helix Propensities of Homo-Oligomers in Aqueous Solution Studied by Multicanonical Algorithm," *Chemical Physics Letters* **309**, 95-100 (1999).

Ayori Mitsutake and Yuko Okamoto, "Helix-Coil Transitions of Amino-Acid Homo-Oligomers in Aqueous Solution Studied by Multicanonical Simulations," submitted.

5.1 Introduction

Helix-coil transitions of peptide and protein systems provide important insight into protein folding problem. There exist extensive recent studies both in experiment [1]-[8] and in theory [9]-[20].

In a previous work,[17] thermodynamics of helix-coil transitions of homo-oligomers in gas phase were studied by multicanonical algorithms [21]. Multicanonical algorithm is particularly useful for studying helix-coil transitions, because a single simulation run can give information about wide range of temperatures covering both helix phase at low temperatures and coil phase at high temperatures. Homo-oligomers of length 10 were considered for three characteristic amino acids, alanine (helix former), valine (helix indifferent), and glycine (helix breaker). It was shown that the obtained helix propagation parameters s of the Zimm-Bragg model [22] for the three amino acids were in remarkable agreement with the experimental values [17]. The helix-coil transition from an ideal helix to a random coil was observed in homo-alanine. It turned out, however, that the transition temperature was above 400 K and unrealistically high [17]. It was conjectured that the lack of solvent effects in the analysis caused this discrepancy.

In this chapter, we study thermodynamics of helix-coil transitions in amino-acid homo-oligomers of length 10 in aqueous solution by multicanonical algorithm and compare the results with those in gas phase. Homo-alanine, homo-valine, and homo-glycine are considered. We calculate average values of total potential energy, its component terms, helicity, and specific heat as a function of temperature. Zimm-Bragg s and σ parameters are also obtained as a function of temperature. The helix-coil transition temperature for homo-alanine in aqueous solution is indeed found to be much lower than that in gas phase.

5.2 Computational Details

In this chapter we employ the multicanonical simulation to estimate various thermodynamic quantities as a function of temperature. The solvation free energy that we used is the sum of terms that are proportional to the solvent-accessible surface area of the atomic

groups of the solute (see Eq. (2.7)).

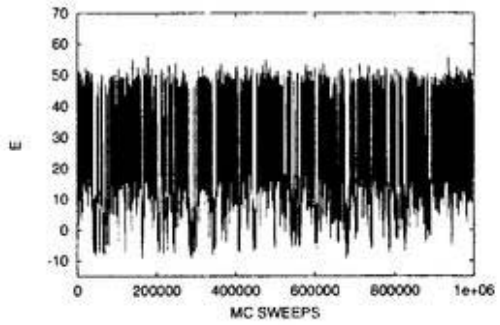
In the present work, the numbers of degrees of freedom (dihedral angles) are 30, 50, and 20 for $(Ala)_{10}$, $(Val)_{10}$, $(Gly)_{10}$, respectively. One Monte Carlo (MC) sweep consists of updating all these angles once with Metropolis evaluation [23] for each update. For the calculation of multicanonical weight factors, we used version 3 of the iterative procedure in subsection 2.2.2; it required between 110 000 MC sweeps in gas phase and 400 000 MC sweeps in aqueous solution for $(Ala)_{10}$, 900 000 MC sweeps in gas phase and 1 000 000 MC sweeps in aqueous solution for $(Val)_{10}$, and 450 000 MC sweeps in gas phase and 250 000 MC sweeps in aqueous solution for $(Gly)_{10}$. After the optimal weighting factor $w_{mu}(E)$ was determined we then made one production run with 1 000 000 MC and obtained various thermodynamic quantities as a function of temperature by the reweighting techniques (see Eqs. (2.32) and (2.33)). Initial conformations were randomly generated.

5.3 Results and Discussion

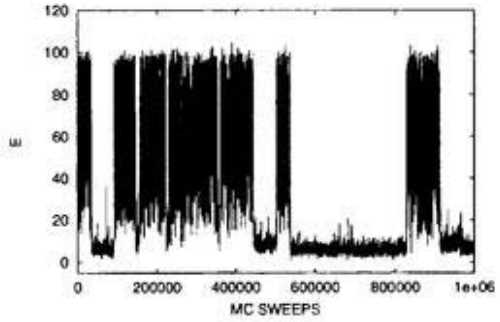
5.3.1 Time Series of the Total Potential Energy

We first examine how much of the configuration space the multicanonical simulations explore. As explained in section 2.2, a simulation in the multicanonical ensemble performs a one-dimensional random walk in the potential energy space. It should visit not only the ground-state regime but also states with very high energy. This ensures that the simulations avoid getting trapped in configurations with energy local minima. We display in Fig. 5.1 the "time series" of the total potential energy E_{TOT} for $(Ala)_{10}$, $(Val)_{10}$, and $(Gly)_{10}$ in aqueous solution. The results indeed exhibit a random walk in energy space covering a range of 60 – 100 kcal/mol. We confirmed that random walks in energy space for $(Ala)_{10}$, $(Val)_{10}$, and $(Gly)_{10}$ were also obtained in gas phase (data not shown). Since configurations are sampled over a large range of energies, the reweighting techniques allow one to calculate thermodynamic quantities as functions of temperature for a wide range of temperatures (see Eqs. (2.32) and (2.33)).

(a)



(b)



(c)

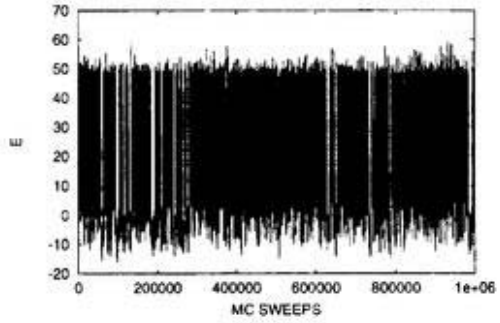


Figure 5.1: Time series of the total potential energy E_{TOT} (kcal/mol) for $(Ala)_{10}$ (a), $(Val)_{10}$ (b), and $(Gly)_{10}$ (c) in aqueous solution.

5.3.2 The Lowest-Energy Conformations

We investigate the lowest-energy conformations of homo-oligomers obtained during the present simulations. In Tables 5.1 and Table 5.2, we list the dihedral angles of the lowest-energy conformations for $(Ala)_{10}$, $(Val)_{10}$, and $(Gly)_{10}$ in gas phase and in aqueous solution, respectively.

For $(Ala)_{10}$ in both environments the dihedral angles from residue 2 to residue 9 are in α-helix state. Here, the criterion we adopt for α-helix state of residue is as follows: We consider that a residue is in the α-helix configuration when the dihedral angles (ϕ, ψ) fall in the range ($-70 \pm 30°$, $-37 \pm 30°$). The length ℓ of a helical segment is then defined by the number of successive residues which are in the helix configuration. The lowest-energy conformations for $(Ala)_{10}$ have the helix length $\ell = 8$ and are completely helical conformations (the terminal residues tend to be frayed). The side-chain structures of $(Ala)_{10}$ is also uniquely determined for the lowest-energy conformations in both environment; namely, the values of χ are close to one of 60°, −60°, and 180°, which are all equivalent angles because of the 3-fold rotational symmetry of the alanine side chain. These lowest-energy conformations for $(Ala)_{10}$ in gas phase and in aqueous solution are shown in Fig. 5.2 (a) and Fig. 5.2 (b), respectively. These conformations have six intra-chain backbone hydrogen bonds that characterize the α-helix and are indeed completely helical.

As shown in Tables 5.1 and 5.2, the dihedral angles of $(Val)_{10}$ for the lowest-energy conformations in both environments are again in almost ideal helix state (from residue 2 to residue 9 in gas phase and from residue 2 to residue 8 in aqueous solution). These lowest-energy conformations are shown in Fig. 5.2 (c) and Fig. 5.2 (d). We remark that the lowest-energy conformation of $(Val)_{10}$ in aqueous solution discussed here was actually obtained by another multicanonical simulation run that was made separately from the one shown in Fig. 5.1 (b) (the energy difference between the lowest-energy conformations obtained by the two runs was only about 1 kcal/mol). This run, however, got trapped in the ground state region and did not preform a random walk in energy space. We thus used the results of the run in Fig. 5.1 (b) instead for the calculation of thermodynamic quantities as a function of temperature (which are presented below). The thermodynamic

Table 5.1: The dihedral angles of the lowest-energy conformations in gas phase for $(Ala)_{10}$ (a), $(Val)_{10}$ (b), and $(Gly)_{10}$ (c).

(a)

residue	ϕ	ψ	χ
1	−24	−42	−68
2	−66	−38	−67
3	−68	−36	65
4	−66	−38	65
5	−69	−35	−57
6	−71	−35	63
7	−73	−30	177
8	−75	−36	51
9	−75	−34	−50
10	−153	97	178

(b)

residue	ϕ	ψ	χ^1	χ^2	χ^3
1	3	−20	65	−51	177
2	−70	−34	173	−67	−59
3	−64	−31	173	172	−54
4	−73	−39	174	−68	60
5	−64	−34	166	47	39
6	−66	−34	163	50	−71
7	−73	−36	163	170	49
8	−63	−34	161	−72	50
9	−78	−47	166	48	168
10	−99	91	−176	178	−57

(c)

residue	ϕ	ψ
1	122	−50
2	160	−86
3	−148	69
4	61	−109
5	−56	−52
6	−79	78
7	146	−34
8	84	31
9	170	174
10	−51	144

Table 5.2: The dihedral angles of the lowest-energy conformations in aqueous solution for $(Ala)_{10}$ (a), $(Val)_{10}$ (b), and $(Gly)_{10}$ (c).

(a)

residue	ϕ	ψ	χ
1	19	149	62
2	−76	−29	−166
3	−68	−38	−175
4	−74	−33	174
5	−69	−39	−46
6	−67	−36	63
7	−68	−40	67
8	−64	−36	69
9	−70	−42	63
10	−154	107	174

(b)

residue	ϕ	ψ	χ^1	χ^2	χ^3
1	13	−32	62	−51	55
2	−62	−33	173	−63	−48
3	−67	−40	171	172	62
4	−68	−35	170	48	173
5	−65	−34	163	−65	−72
6	−77	−39	165	−72	−72
7	−60	−40	162	−73	42
8	−79	−47	170	−72	55
9	−106	84	−172	−63	71
10	−79	128	175	−61	57

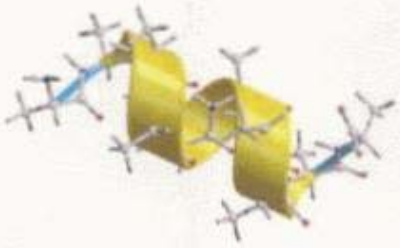
(c)

residue	ϕ	ψ
1	107	−83
2	79	−89
3	170	−46
4	−78	173
5	61	35
6	62	35
7	80	44
8	−152	29
9	176	172
10	−69	−79

quantities were calculated for the temperature range between 200 K and 700 K, which the ground state corresponds to the temperature 0 K. Thus, it does not matter whether the run did reach the ground state or not, as long as it performs a random walk in the relevant energy range.

For $(Gly)_{10}$ in gas phase, the dihedral angles of the lowest-energy conformation seem to imply that it is a coil structure (see Table 5.1). The lowest-energy conformation of $(Gly)_{10}$ in aqueous solution, on the other hand, has a left-handed α-helix from residue 5 to residue 7 (see Table 5.2). Both conformations are compared in Fig. 5.2 (e) and 5.2 (f). It is apparent that they are rather compact and round. A close examination of the structures revealed that both conformations have β-sheet-like characteristics. Namely, they are stabilized by the intrachain backbone hydrogen bonds that are found in β-sheet structures. For the lowest-energy conformation in gas phase (Fig. 5.2 (e)), the carbonyl oxygen (and amide nitrogen) of residue 3 and amide nitrogen (and carbonyl oxygen) of residue 6 are hydrogen bonded. There also exist two such hydrogen bonds that connect residues 4 and 9. For the lowest-energy conformation in aqueous solution (Fig. 5.2 (f)), there exist two such hydrogen bonds between residue 2 and residue 9.

(a)



(b)

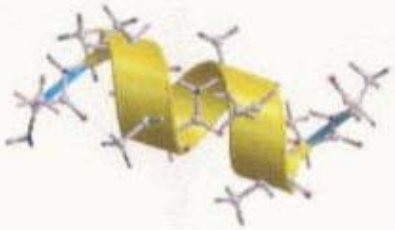
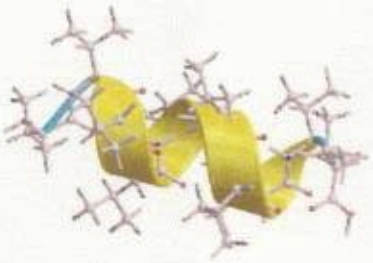


Figure 5.2: The lowest-energy conformations of $(Ala)_{10}$ ((a) and (b)), $(Val)_{10}$ ((c) and (d)), and $(Gly)_{10}$ ((e) and (f)) in gas phase and in aqueous solution, respectively.

(c)



(d)

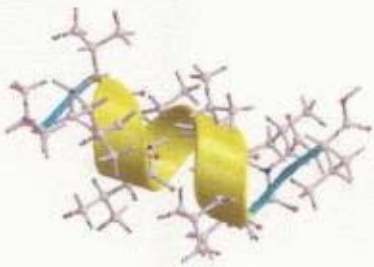


Fig. 5.2 (Continued)

(e)

(f)

Fig. 5.2 (Continued)

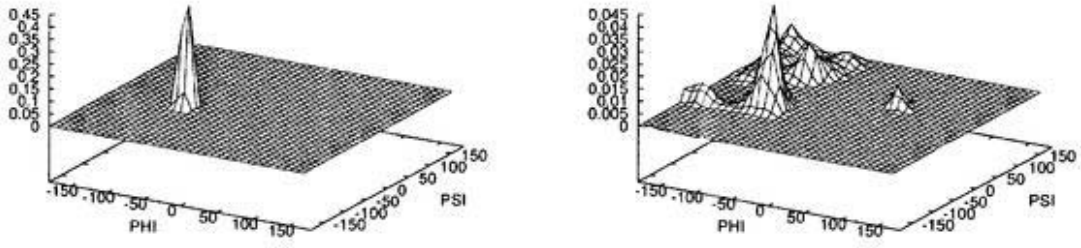
5.3.3 Distributions of Backbone Dihedral Angles

The lowest-energy conformations discussed in detail in the last subsection correspond to the conformations at low temperatures. We next calculated the distributions of the backbone dihedral angles of homo-oligomers in aqueous solution as a function of temperature in order to study how these conformations change as the temperature is raised. In Fig. 5.3 we show the distributions of the backbone dihedral angles of the sixth residue of $(Ala)_{10}$, $(Val)_{10}$, and $(Gly)_{10}$ in aqueous solution at $T = 200$ K and 1000 K. For $(Ala)_{10}$ there is a single peak at $T = 200$ K and this peak corresponds to the dihedral angle of a right-handed α-helix state. The distributions for other residues have essentially the same peak except for the terminal residues. This implies that around $T = 200$ K there exists only a completely helical conformation. At $T = 1000$ K the distributions are widely spread, implying the large thermal fluctuations. These results suggest the existence of a transition between an ensemble of well-defined compact conformations (ideal α-helix state) and random-coil structures.

For $(Val)_{10}$ (see Fig. 5.3 (b)), the results are similar to those for $(Ala)_{10}$ in the sense that at $T = 200$ K there is a dominant peak in the distribution that corresponds to the lowest-energy conformation and that at $T = 1000$ K the distribution is widely spread (random-coil state).

The situation is slightly different for $(Gly)_{10}$. Since glycine dose not have a side chain, $(Gly)_{10}$ is much more flexible than the other two homo-oligomers. We observe two dominant peaks in the distributions at $T = 200$ K (see Fig. 5.3 (c)), which implies that this temperature is not low enough to single out the lowest-energy conformation. The large flexibility of $(Gly)_{10}$ is most clearly seen in the distributions at $T = 1000$ K.

(a)



(b)

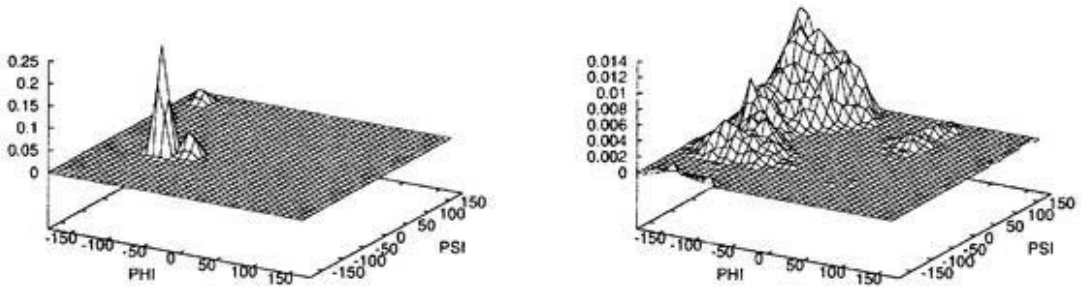


Figure 5.3: Distributions of the backbone dihedral angles of $(\mathrm{Ala})_{10}$ (a), $(\mathrm{Val})_{10}$ (b), and $(\mathrm{Gly})_{10}$ (c) in aqueous solution as a function of temperature. The results for the sixth residue from the N-terminus at $T = 200$ K (left-hand side) and 1000 K (right-hand side) are shown. The values for each case were calculated from one multicanonical production run of 1 000 000 MC sweeps.

(c)

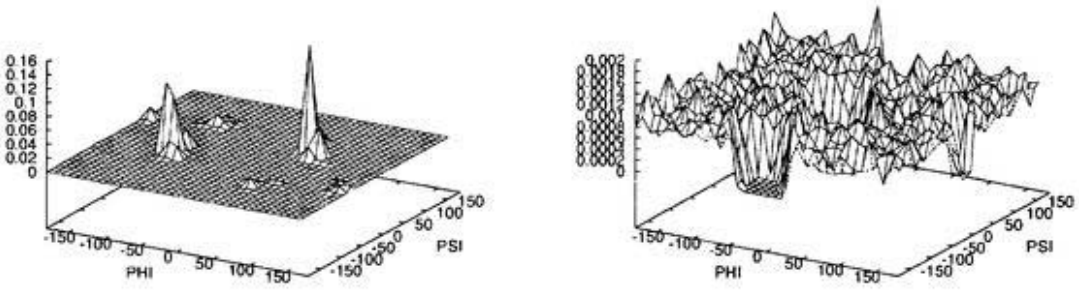


Fig. 5.3 (Continued)

5.3.4 Average Energy and Specific Heat

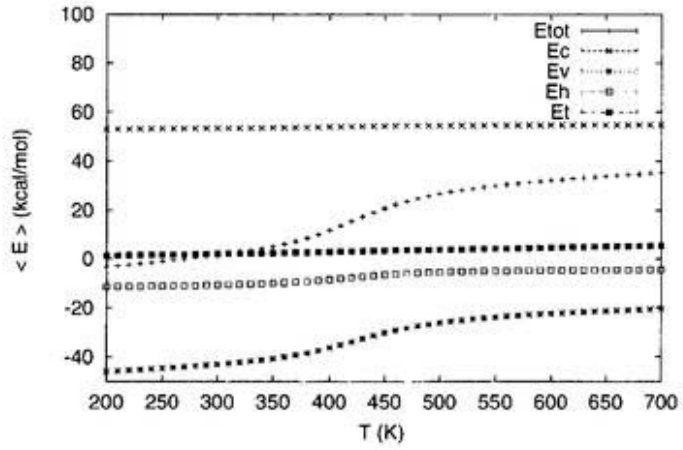
We investigate how each energy term varies as a function of temperature. We calculated the average values of total energy and its component terms of the homo-oligomers, $(Ala)_{10}$, $(Val)_{10}$, and $(Gly)_{10}$, as a function of temperature in gas phase and in aqueous solution. In Fig. 5.4 the results are shown. For homo-alanine in gas phase, all the conformational energy terms increase monotonically as temperature increases. The changes of each component terms are very small except for the Lennard-Jones term, E_V, indicating that E_V plays an important role in the folding of homo-alanine [17]. The Lennard-Jones term, in principle, has contributions from all possible pairs of atoms, while hydrogen-bond term is only from the donor-acceptor pairs. Therefore, the Lennard-Jones term is responsible for the large conformational changes from a random coil to an ideal helix. Once the (near-)helical conformation is obtained, the hydrogen-bond term further stabilizes it.

In aqueous solution the overall behaviors of the conformational energy terms are very similar to those in gas phase. The solvation term, on the other hand, decreases monotonically as temperature increases. These results imply that the solvation term favors random-coil conformations, while the conformational terms favor helical conformations. This is because the solvation free energy of conformations at high temperatures (random coil) is lower than that at low temperatures (α-helix conformations) and because the conformational energies at high temperatures (random coil) are higher than those at low temperatures (α-helix conformations). The rapid changes (decrease for the solvation term and increase for the rest of the terms) of all the average values occur at the same temperature (around at 420 K in gas phase and 340 K in solvent). This suggests the existence of a certain phase transition. As was shown in gas phase in Ref. [17] and is discussed below for the case with solvent, this transition indeed corresponds to a helix-coil transition. It is interesting to note that the helix-coil transition in solvent is the result of two conflicting effects between conformational energy and solvation free energy, which lowers the helix-coil transition temperature compare to the gas-phase value.

In Fig. 5.5 the average energy values as a function of temperature for $(Val)_{10}$ and $(Gly)_{10}$ in aqueous solution are shown. For homo-valine and homo-glycine, the behaviors of the conformational energy terms in aqueous solution are quite similar to those in

gas phase (data not shown) and all the conformational terms increase monotonically as temperature increases. The solvation term, on the other hand, decreases monotonically as a function of temperature. For homo-valine and homo-glycine, the change in total energy is not as conspicuous as in homo-alanine. Hence, the helix-coil transition in homo-valine and homo-glycine is not as clearly observed as in homo-alanine.

(a)



(b)

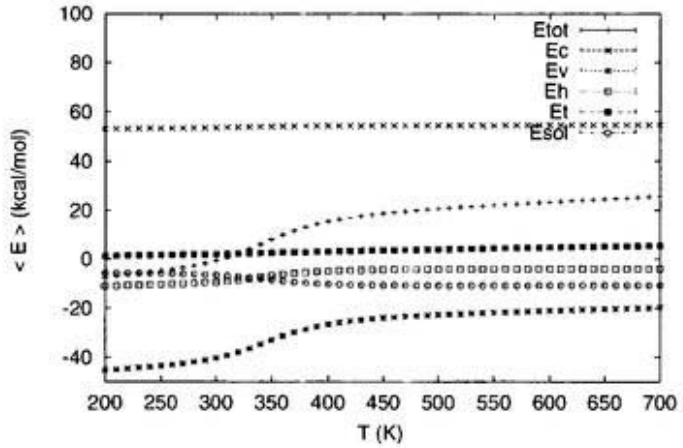
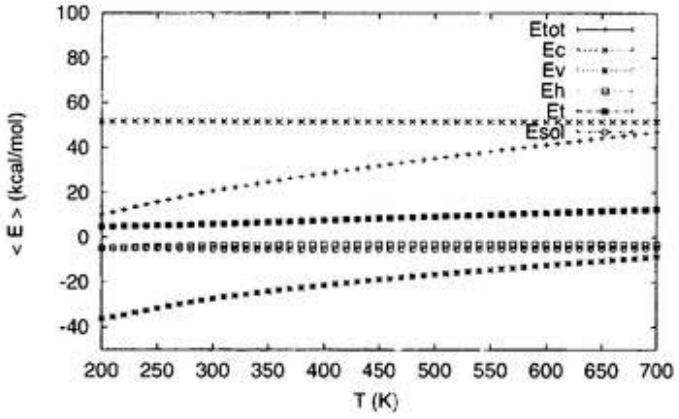


Figure 5.4: Average total energy E_{TOT} and averages of its component terms, Coulomb energy E_C, hydrogen-bond energy E_H, Lennard-Jones energy E_V, torsion energy E_T, and solvation free energy E_{SOL} (only for the case in aqueous solution) for homo-alanine as a function of temperature T in gas phase (a) and in aqueous solution (b). The values for each case were calculated from one multicanonical production run of 1 000 000 MC sweeps.

(a)



(b)

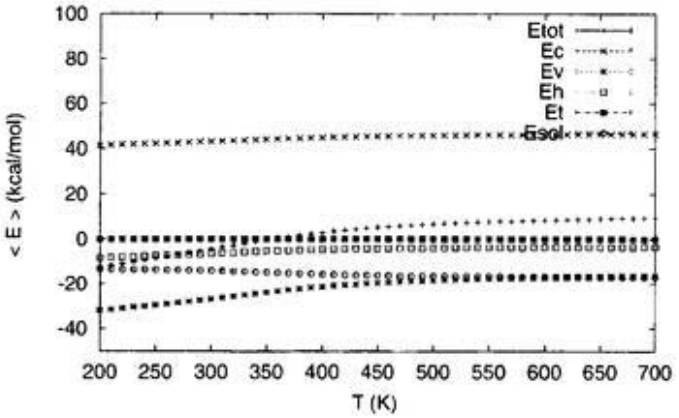


Figure 5.5: Average total energy E_{TOT} and averages of its component terms, Coulomb energy E_C, hydrogen-bond energy E_H, Lennard-Jones energy E_V, torsion energy E_T, and solvation free energy E_{SOL} (only for the case in aqueous solution) for homo-valine (a) and home-glycine (b) as a function of temperature T in aqueous solution. The values for each case were calculated from one multicanonical production run of 1 000 000 MC sweeps.

We calculated the specific heat for the homo-oligomers as a function of temperature. The specific heat here is defined by the following equation:

$$C(T) = \beta^2 \frac{< E_{TOT}^2 >_T - < E_{TOT} >_T^2}{N}, \tag{5.1}$$

where N (= 10) is the number of residues in the oligomer. In Fig. 5.6 we show the specific heat as a function of temperature for $(\mathrm{Ala})_{10}$ in gas phase and in aqueous solution. We

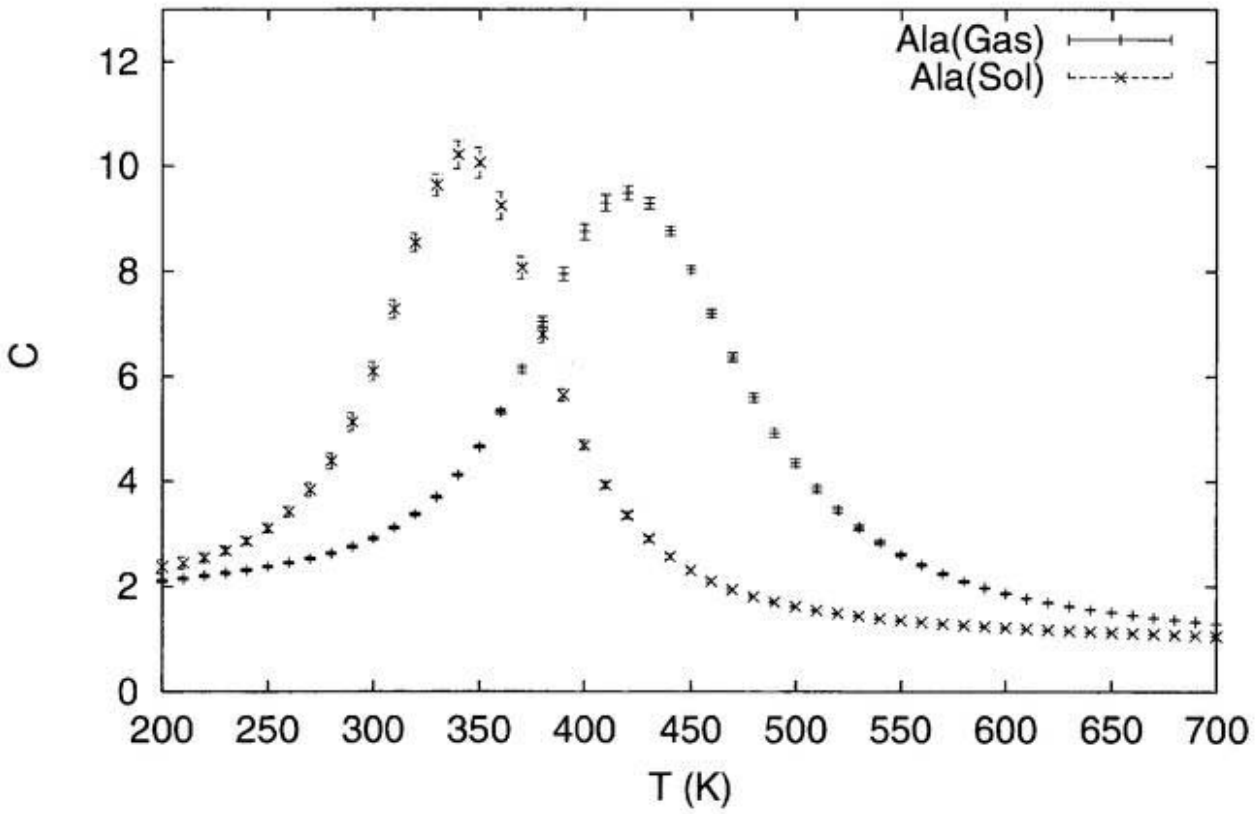


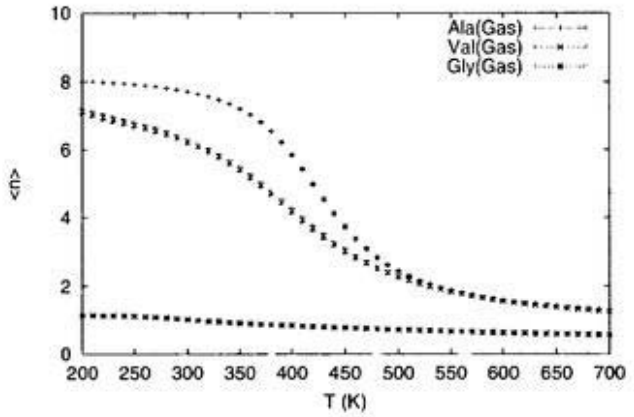
Figure 5.6: Specific heat C as a function of temperature T for $(\mathrm{Ala})_{10}$ in gas phase and in aqueous solution. The values for each case were calculated from one multicanonical production run of 1 000 000 MC sweeps.

observe sharp peaks in the specific heat for both environment, which implies the existence of some phase transition. The temperatures at the peak, transition temperatures, are $T_c \approx 420$ K and 340 K in gas phase and in aqueous solution, respectively. The transition temperature T_c for $(\mathrm{Ala})_{10}$ in aqueous solution is thus significantly lower than that in gas phase and much closer to experimentally relevant temperatures.

5.3.5 Helicity of Homo-Oligomers

We calculated the average number of helical residues $< n >_T$ in a conformation as a function of temperature. In Fig. 5.7 we show the average helicity $< n >_T$ as a function of temperature for $(\mathrm{Ala})_{10}$, $(\mathrm{Val})_{10}$, and $(\mathrm{Gly})_{10}$ in gas phase and in aqueous solution. The

(a)



(b)

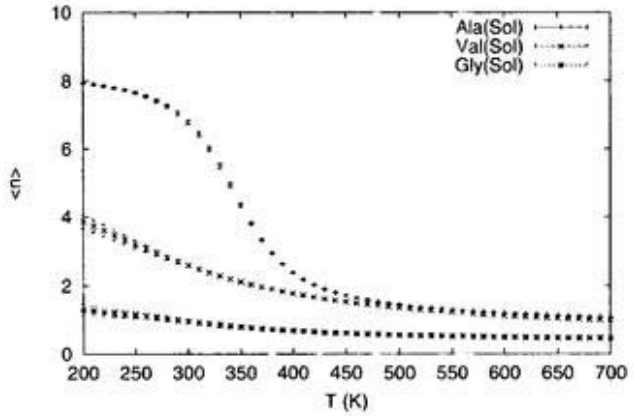


Figure 5.7: Average helicity $< n >_T$ as a function of temperature T for $(\mathrm{Ala})_{10}$, $(\mathrm{Val})_{10}$, and $(\mathrm{Gly})_{10}$ in gas phase (a) and in aqueous solution (b). The values for each case were calculated from one multicanonical production run of 1 000 000 MC sweeps.

average helicity tends to decrease monotonically as the temperature increases because of the increased thermal fluctuations.

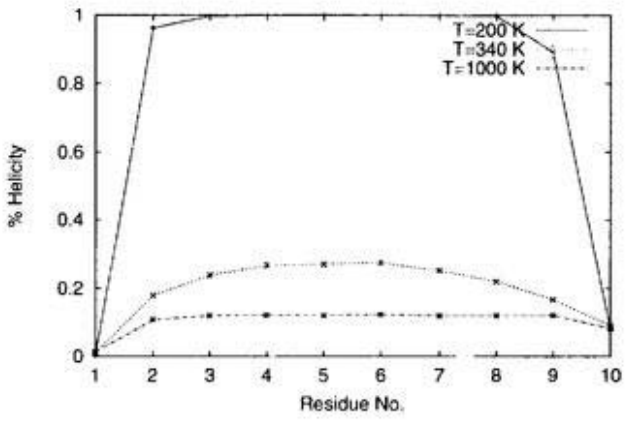
At $T = 200$ K, $< n >_T$ for homo-alanine in both environment are 8. If we neglect the terminal residues, in which α-helix tends to be frayed, $n = 8$ corresponds to the maximal helicity, and the conformation can be considered completely helical. Then homo-alanine

in both environment is in an ideal helical structure at $T = 200$ K. Around the room temperature, $(Ala)_{10}$ in gas phase and in aqueous solution is still substantially helical ($\approx$ 80 and $\approx$ 70 % helicity), respectively. This is consistent with the fact that alanine is a strong helix former. For homo-alanine, at transition temperatures (around 420 K in gas phase and 340 K in aqueous solution), $< n >_T$ is 5 (50 % helicity). This implies that the phase transitions observed above by the peak in specific heat are indeed a helix-coil transition between an ideal helix and a random coil. Hence, as far as the helix-coil transition is concerned, the solvation effects do not alter the nature of the transition and just shift the transition temperature.

As is shown in Fig. 5.7, the average helicity of homo-glycine in aqueous solution is similar to that in gas phase and is very low ($<$ 20 % helicity). It is apparent that homo-glycine does not favor helix formation over the whole temperature range in both environment. The average helicity of homo-valine in aqueous solution is lower than that in gas phase and is less than 40 % helicity for a wide range of temperatures. The percent helicity lies between that of alanine and glycine. All these results are consistent with the fact that alanine is a helix former and glycine is a helix breaker, while valine comes in between the two.

We next calculated the percent helicity as a function of residue number for the three homo-oligomers. The results at $T = 200$, 340, and 1000 K are shown in Fig. 5.8. The percent helicity is in general lower at the terminal residues, than in the internal residues (i.e., fraying is observed) for a wide range of temperatures, because the dihedral angles of terminal residues are less constrained than those of the internal residues. For $(Ala)_{10}$ and $(Val)_{10}$, the internal residues are significantly helical at $T = 200$ K. For $(Gly)_{10}$ the residues prefer a coil state rather than helix for a wide range of temperatures. We do observe fraying of for all cases. The contrast is most outstanding for $(Ala)_{10}$ because it has high helicity. The increase of fraying as the temperature is raised is also clearly seen for $(Ala)_{10}$.

(a)



(b)

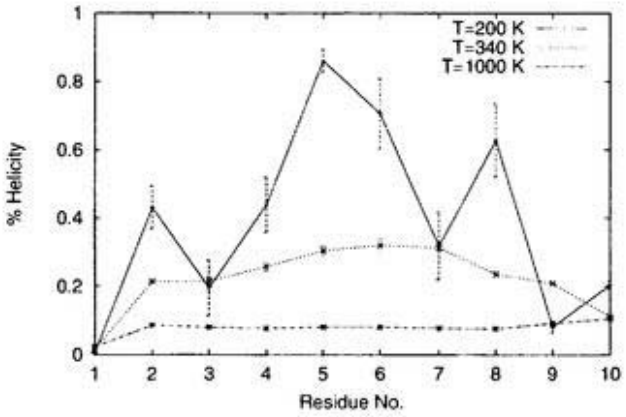


Figure 5.8: Percent helicity as a function of residue number for $(\mathrm{Ala})_{10}$ (a), $(\mathrm{Val})_{10}$ (b), and $(\mathrm{Gly})_{10}$ (c) in aqueous solution at $T = 200$, 340, and 1000 K. The values for each case were calculated from one multicanonical production run of 1 000 000 MC sweeps.

(c)

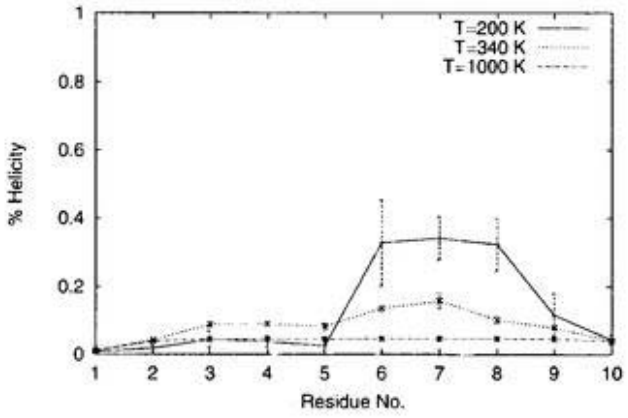


Fig. 5.8 (Continued)

5.3.6 Zimm-Bragg s and σ Parameters

Finally, the helix propagation parameter s and nucleation parameter σ of Zimm-Bragg model [22] were calculated as a function of temperature.

According to the Zimm-Bragg model, the average number of helical residues $< n >_T$ and the average length $< \ell >_T$ of a helical segment are given for large N by

$$\begin{aligned} \frac{< n >_T}{N} &= \frac{1}{2} - \frac{1-s}{2\sqrt{(1-s)^2 + 4s\sigma}}, \\ < \ell >_T &= 1 + \frac{2s}{1-s+\sqrt{(1-s)^2 + 4s\sigma}}, \end{aligned} \tag{5.2}$$

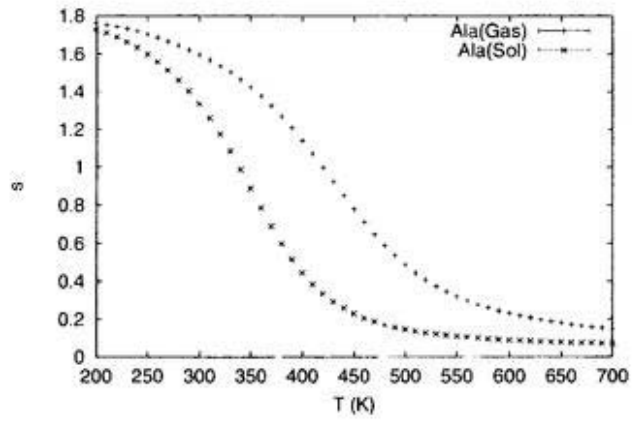
where N is the number of residues. Note that from Eq. (5.2) the temperature where $s = 1$ holds corresponds to 50 % helicity, which in turn gives the helix-coil transition temperature. From these equations with the values of $\frac{<n>}{N}$ and $< \ell >$ calculated from the multicanonical production runs, one can obtain estimates of s and σ parameters.

In Fig. 5.9 we show the s and σ values for alanine as a function of temperature. The s parameter monotonically decrease as the temperature increases. It is shown that s(Ala) in aqueous solution decreases more rapidly than that in gas phase as the temperature increases. As noted above, the helix-coil transition temperature T_c can also be identified as the temperature where s =1 holds (i.e., 50 % helicity) in Fig. 5.9 (a). It is 420 K in gas phase and 340 K in aqueous solution in agreement with the previous estimates by the peaks of specific heat (see Fig. 5.6).

As is clear from Fig. 5.9 (b), in the helix phase ($T < T_c$) the σ parameter for alanine is small and constant, but in the random-coil phase ($T > T_c$) σ starts to grow as temperature increases. This growth of σ value reflects the increased thermal fluctuations that prevent the formation of a long helix. That is, below T_c cooperatively for helix formation wins over thermal fluctuations, but above T_c thermal fluctuations win and no long helices can be formed.

We next make a comparison between the s values obtained by the present simulations and those by experiments [1]-[8]. In Table 5.3 we list s for the three amino acids at T = 273 K in gas phase and in aqueous solution obtained by simulations together with those by experiments. The values in aqueous solution are slightly smaller than those in gas

(a)



(b)

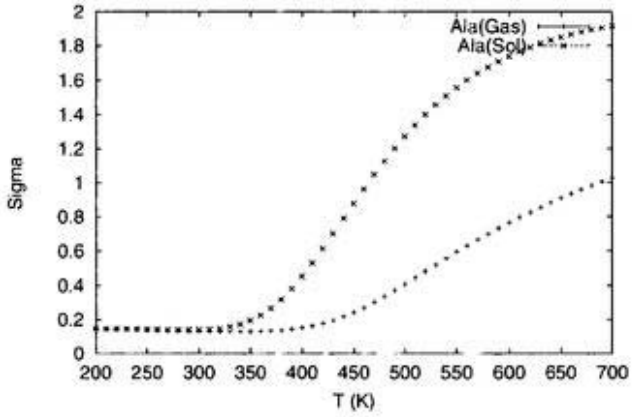


Figure 5.9: Helix propagation parameter s (a) and nucleation parameter σ (b) of the Zimm-Bragg model as a function of temperature T for alanine in gas phase and in aqueous solution. The values for each case were calculated from one multicanonical production run of 1 000 000 MC sweeps.

phase but not too much different except for valine. The s value for valine is significantly lower in aqueous solution than in gas phase. One finds that our values in aqueous solution are in remarkable agreement with those determined by experiments.

Table 5.3: The helix propagation parameter s and nucleation parameter σ of Zimm-Bragg model for alanine (Ala), valine (Val), and glycine (Gly) at $T = 273$ K in gas phase (Gas) and in aqueous solution (Sol) together with the experimental values (Exp).

Amino acid	s (Gas)	s (Sol)	s (Exp)
Ala	1.67	1.51	$1.5 \sim 2.19$
Val	1.36	0.31	$0.2 \sim 0.93$
Gly	0.17	0.21	$0.05 \sim 0.57$

Bibliography

[1] S. Padmanabhan, S. Marqusee, T. Ridgeway, T. M. Laue, and R. L. Baldwin, *Nature* **344**, 268 (1990).

[2] J. Wojcik, K. H. Altman, and H. A. Scheraga, *Biopolymers* **30**, 121 (1990).

[3] G. Merutka, W. Lipton, W. Shalango, S. H. Park, and E. Stellwagen, *Biochemistry* **29**, 121 (1990).

[4] K. T. O'Neil and W. F. DeGrado, *Science* **250**, 646 (1990).

[5] P. C. Lyu, M. I. Liff, L. A. Marky, and N. R. Kallenback, *Science* **250**, 669 (1990).

[6] A. Horovitz, J. M. Matthews, and A. R. Fersht, *J. Mol. Biol.* **227**, 560 (1992).

[7] A. Chakrabartty, T. Kortemme, and R. L. Baldwin, *Protein Sci.* **3**, 843 (1994).

[8] J. K. Myers, C. N. Pace, and J. M. Scholtz, *Proc. Natl. Acad. Sci. USA* **94**, 2833 (1997).

[9] M. Fukugita, T. Nakazawa, H. Kawai, and Y. Okamoto, *Chem. Lett.* **1991**, 1279.

[10] T. Ooi and M. Oobatake, *Proc. Natl. Acad. Sci. USA* **88**, 2859 (1991).

[11] J. Tirado-Rives and W. L. Jorgensen, *Biochemistry* **30**, 3864 (1991).

[12] D. J. Tobias and C. L. Brooks, III, *Biochemistry* **30**, 6059 (1991).

[13] T. P. Creamer and G. D. Rose, *Proc. Natl. Acad. Sci. USA* **89**, 5937 (1992).

[14] V. Daggett and M. Levitt, *J. Mol. Biol.* **223**, 1121 (1992).

[15] Y. Okamoto, *PROTEINS: Struct. Funct. Genet.* **19**, 14 (1994).

[16] Y. Okamoto, *Biopolymers* **34**, 529 (1994).

[17] Y. Okamoto and U. H. E. Hansmann, *J. Phys. Chem.* **99**, 11276 (1995).

[18] M. Takano, T. Takahashi, and K. Nagayama, *Phys. Rev. Lett.* **80**, 5691 (1998).

[19] U. H. E. Hansmann and Y. Okamoto, *J. Chem. Phys.* **110**, 1267 (1999).

[20] M. Takano, T. Yamato, J. Higo, A. Suyama, and K. Nagayama, *J. Am. Chem. Soc.* **121**, 605 (1999).

[21] B. A. Berg and T. Neuhaus, *Phys. Lett.* **B267**, 249 (1991).

[22] B. H. Zimm and J. K. Bragg, *J. Chem. Phys.* **31**, 526 (1959).

[23] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, *J. Chem. Phys.* **21**, 1087 (1953).

Chapter 6

Conclusions

There are two difficulties that one has to overcome in the protein folding problem. One is how to give the accurate energy function of the system. Especially, the effects of solvent surrounding a protein molecule are non-trivial to incorporate. The other is that there exist a huge number of local minima in the potential energy function, which renders the range of the configuration space that can be sampled by conventional simulation methods very narrow.

In this thesis we explored the ways to alleviate the above difficulties. As for solvation theory, we tried to incorporate a term that is based on the extended scaled particle theory [1, 2] and a term that is proportional to the solvent-accessible surface area of the solute molecule [3]. As for the simulation methods, we employed the Monte Carlo simulated annealing method [4] and multicanonical algorithm [5]. We tested the effectiveness of a few combinations of the above methods with a penta peptide, Met-enkephalin, and three amino-acid homo-oligomers of length 10, homo-alanine, homo-valine, and homo-glycine.

We first tried to combine the extended scaled particle theory and the Monte Carlo simulated annealing method (Chapter 3). The extended scaled particle theory gives a rigorous cavity-formation free energy of solvent. This was the first attempt to combine the extended scaled particle theory and Monte Carlo simulated annealing. Low-energy conformations of Met-enkephalin were successfully classified into several groups of similar structures. The backbone of the lowest-energy conformation in solvent has a shape similar to that of the second lowest-energy structure in gas phase; the global-minimum-energy conformation in gas phase had much higher energy in the simulations in solvent. It seems that the solvent effects decrease the number of low-energy local minima. More-

over, we studied in detail the relations between the solvent-accessible surface area and the cavity-formation free energy. It was shown that the cavity-formation free energy of the obtained conformations is not necessarily proportional to the accessible surface area for a small peptide such as Met-enkephalin. We remark that the lowest-energy conformation of Met-enkephalin obtained by simulations with a rigorous solvation theory (RISM) is fully extended in agreement with the results of NMR experiments [6]. However, the lowest-energy conformation obtained by the present simulations using the extended scaled particle theory is rather round, although some of the low-energy conformations have extended backbone structures. Hence, we find that the neglected contributions (especially, the electrostatic interactions between solute and solvent molecules) are also important when we try to compare the simulation results with experiments. The next thing one should do is probably to combine the rigorous solvation theory and multicanonical algorithm. The work is in progress but it is computationally very demanding, and we tried a simplified solvation theory instead in this thesis.

We performed a multicanonical simulation of Met-enkephalin in gas phase (Chapter 4). We obtained the distributions of hydrogen bonds in backbone. We used the patterns of hydrogen-bond formations to classify conformations into groups of similar structures. It was found that there are at least four well-defined groups which correspond to local-minimum-energy states. The global-minimum-energy state forms a type II' β-turn, the second-lowest-energy state forms a type II β-turn, and the third and fourth states form γ-turns. While with simulated annealing we must repeat many simulations from different initial states to obtain information near the lowest-energy state, multicanonical algorithm allows us to calculate various thermodynamic quantities as a function of temperature for a wide temperature range from a single simulation run. The results such as distributions of conformations as a function of temperature give essential information about the free energy landscape of the peptide. Hence, multicanonical algorithm is a more effective simulation method than the simulated annealing method.

Finally, we performed multicanonical Monte Carlo simulations to study helix-coil transitions of homo-oligomers in aqueous solution based on the solvent-accessible surface area (Chapter 5). Homo-oligomers of length 10 were considered for three characteristic amino

acids, alanine (helix former), valine (helix indifferent) and glycine (helix breaker). Various thermodynamic quantities as a function of temperature were calculated and compared with those obtained in gas phase. It was found that solvation effects reduce helix formation slightly (in fact, it was shown that random-coil conformations are most favored by the solvent), but that overall nature of the helix-coil transition is unaltered by the addition of solvent. For instance, for homo-alanine we still observed in solvent a helix-coil transition from a completely helical conformation at low temperatures to random-coil conformations at high temperatures. This agreement can presumably be accounted for by the fact that we dealt with only nonpolar amino acids, with which we can minimize the complications of electrostatic interactions between the homo-oligomer (side chains) and solvent. It was also shown that the helix-coil transition temperature for homo-alanine gets significantly lowered in aqueous solution compared with that in gas phase. This indeed rectified the unrealistically high value in gas phase, which was the most serious discrepancy between theory and experiments found in the previous work [7].

Bibliography

[1] M. Irisa, T. Takahashi, F. Hirata, and T. Yanagida, *J. Mol. Liquids.* **65**, 381 (1995).

[2] M. Irisa, T. Takahashi, K. Nagayama, and F. Hirata, *Mol. Phys.* **85**, 1227 (1995).

[3] T. Ooi, M. Oobatake, G. Némethy, H. A. Scheraga, *Proc. Natl. Acad. Sci. USA* **84**, 3086 (1987).

[4] S. Kirkpatrick, C. D. Gelatt, Jr., and M. P. Vecchi, *Science* **220**, 671 (1983).

[5] a) B. A. Berg and T. Neuhaus, *Phys. Lett.* **B267**, 249 (1991).
b) B. A. Berg and T. Neuhaus, *Phys. Rev. Lett.* **68**, 9 (1992).

[6] M. Kinoshita, Y. Okamoto, and F. Hirata, *J. Am. Chem. Soc.* **120**, 1855 (1998).

[7] a) Y. Okamoto, U. H. E. Hansmann, and T. Nakazawa, *Chem. Lett.* **1995**, 391.
b) Y. Okamoto and U. H. E. Hansmann, *J. Phys. Chem.* **99**, 11276 (1995).