ELSEVIER

Biophysical Chemistry 65 (1997) 109–121

Biophysical Chemistry

Finite one-dimensional spin systems as models of biopolymers

Takeshi Kikuchi

International Research Laboratories, Ciba-Geigy (Japan) Ltd., 10-66 Miyuki-cho, Takarazuka 665, Japan

Received 6 June 1996; revised 24 June 1996; accepted 21 November 1996

Abstract

New models are proposed for describing various properties of biopolymers, especially those of proteins and nucleic acids. Each model is constituted of a set of spins arranged on a chain, and each pair of spins produces an interaction. We examine the transitions of these spin systems between the ground state and the disordered state. It is found that the transitions of the present spin systems demonstrate various properties in response to values of the so-called interaction energy. If we define interaction energy parameters with no so-called frustration, the system exhibits two-state transitions, similar to the folding–unfolding transition of small proteins. The addition of frustrations to the model produces effects similar to those of mutations in proteins. On the other hand, if the interactions between two spins attenuate as a function of their separation along the chain, the transition of the system has characteristics similar to those of nucleic acids. Thus, the present spin systems can offer a unified view of the folding–unfolding transition of biopolymers in terms of differences in the pairwise interactions between spins. Based on our models, we propose a condition for two-state transition behavior for proteins. © 1997 Elsevier Science B.V.

Keywords: Biopolymers; Protein folding; Spin systems; Monte Carlo; Simulation; Frustration; Two-state transition

1. Introduction

Numerous functions in biological organisms are ultimately attributed to the behavior of two kinds of biopolymers, namely proteins and nucleic acids. For the expression of biological functions of these biopolymers, their tertiary structures and dynamics are quite significant. Above all, the remarkable characteristics of the proteins are the uniqueness of their tertiary structures and the approximate two-state folding–unfolding transition under biological conditions [1–3]. On the other hand, most nucleic acids show a rather loose folding–unfolding transition [4], although tRNA folds into its unique native structure under biological conditions [5]. Thus, forming a unique tertiary structure is not limited to proteins. However, ordinary polymers would fold into compact random structures under poor solvent conditions. The question is asked as to what factors account for these differences among polymers.

Much work has been carried out on the theoretical elucidation of the behavior of proteins [6,7]. Some models based on polymers in a lattice succeed in reproducing the basic properties of protein folding [8–18]. Early work on a lattice model of proteins by Go and co-workers [18] has already revealed that the two-state transition is driven by long- range interactions between residues. It was recently demonstrated by means of a lattice model that the uniqueness of the native structures and the short folding time of the

PII S0301-4622(96)02261-2

proteins are guaranteed by a sufficiently low energy of the ground state compared with the first conformational excited state [16,17]. Various models of the two-state behavior of proteins were also proposed. Shakhnovich and Finkelstein [19,20] proposed that the two-state transition be ascribed to entropy change by side-chain motion. Hao and Scheraga [21,22] also constructed a lattice model which exhibits two-state coil–globule transitions. However, as Lattman and Rose [23] recently pointed out, even if the native state of a protein is energetically destabilized by the mutations of the residues to a fair degree, the two-state folding–unfolding transition is still maintained, and the essential part of the original native structure is conserved. They emphasized that the crucial factor for the protein properties is not only structural stability, but also the specificity of internal interactions. Nishikawa [24] also pointed out that the tertiary structure of a protein does not drastically change by mutation of the residues. A theory to describe the behavior of proteins should explain this property.

Recently, an analogy of the property of proteins to that of a spin glass attracted the attention of many research workers. Stein [25] derived a state density function of a protein by the application of a random energy model to the protein problem. He compared the results of the calculations with his formula for a hemoglobin–CO recombination curve, which seems to reflect the distribution of the conformational substates of a protein, and found good coincidence between them. Bryngelson and Wolynes [26] also assumed random energy distribution in a protein, proposed a phase diagram for an internal interaction parameter and transition temperature, and estimated the average first passage time of protein folding [27]. Recently, Wolynes and co-workers [38,39] analyzed the energy landscape of protein folding using a lattice model based on the random energy model. Shakhnovich and Gutin [28] considered a wide internal interaction distribution as a condition for folding into a unique (or small number of) structure(s) based on the analogy between protein properties and Parisi's solution in spin glass theory. Moreover, they proposed conditions for the folding–unfolding transition of proteins [29,30]. In these theories, the main interest was in the folding of a protein into a unique structure within a reasonable time, but they did not pay sufficient attention to the two-state behavior of the transition. However, it is a quite interesting fact that the behavior of spin systems resembles that of proteins.

On the other hand, the loose melting of nucleic acids can be described basically by the one-dimensional Ising model [31,32]. In the same way, the thermal transition of homopolypeptides can be treated by the one-dimensional Ising model [33,34], i.e. the actions of biopolymers can be modeled well by spin systems.

We propose four models, constituted of spin systems, that can reproduce the basic properties of biopolymers. An advantage of using such simple models is that we can take a broad view of the actions of the biomolecules. The present study focuses mainly on the condition for the two-state transition. We propose spin systems as models of the biopolymers and analyze the overall characteristics of the systems in Section 2. The method used for the numerical calculations is described in Section 3, and these results are presented in Section 4. Finally, we discuss the properties of our models in regard to the folding–unfolding nature of real biopolymers in Section 5.

2. Models

We consider a system consisting of N particles arranged in a linear array. Each particle models each monomer in a biopolymer and can take two spin states. The Hamiltonian of the system is defined as follows

$$E = \sum_{i=1}^{N} \eta_i \sigma_i + \sum_{i>j} J_{ij} \sigma_i \sigma_j + \sum_{\alpha=3}^{N-1} \sum_{i=1}^{N-\alpha+1} K_i^{\alpha} \prod_{\kappa=i}^{i+\alpha-1} \sigma_\kappa \qquad (1)$$

Each σ_i implies the spin state of the ith particle and takes a value of either $+1$ or -1 as the usual Ising model. The σ_i can be regarded as a simplified expression of a conformation of the ith monomer of a polymer. The values of η_i, J_{ij} and K_i^{α} are energy parameters which, in principle, can take arbitrary values. The second term of Eq. (1) contains the coupling of every pair of particles. Therefore, the first and second terms of Eq. (1) do not denote

one-dimensionality of this system. The third term contains every possible correlation between the consecutive particles. The α represents the number of consecutive particles. Only the correlation of the spins of the particles along the chain are taken into account among all the more than three body correlations. One-dimensionality is brought into this system by the third term. In a polymer system, the interaction between or proximity of two monomers is strongly affected by the conformations of the intervening part between these sites. The third term in Eq. (1) corresponds to this effect.

According to the methods of selection of the values of η_i, J_{ij}, and K_i^{α}, we define the following models.

2.1. Random system

Here $\eta_i = a\eta_{i'}$, $J_{ij} = bJ_{ij'}$ and $K_i^{\alpha} = bK_i^{\alpha'}$, where $\eta_{i'}$, $J_{ij'}$ and $K_i^{\alpha'}$ are taken as homogeneous random numbers in the interval $\{-1,1\}$, and a and b are arbitrary positive coefficients.

2.2. No-frustration system

The values of η_i are basically taken in the same way as those of the random system. However, for the values of J_{ij} and K_i^{α}, the absolute values of the parameters, $|J_{ij'}|$ and $|K_i^{\alpha'}|$, are randomly determined within the interval $\{0,1\}$. To avoid a frustration in the system, the signs of $J_{ij'}$ and $K_i^{\alpha'}$ are determined as follows.

$$\mathrm{Sgn}(J_{ij'}) = \mathrm{Sgn}(\eta_{i'})\,\mathrm{Sgn}(\eta_{j'})$$

$$\mathrm{Sgn}(K_i^{\alpha'}) = \prod_{k=i}^{i+\alpha-1} \mathrm{Sgn}(\eta_k) \qquad (2)$$

The no-frustration system corresponds to the principle of the consistency of protein structures proposed by Go [6] and to the minimum frustration model of proteins of Bryngelson and Wolynes [26].

2.3. Partly frustrated system

Frustrations are introduced into the no-frustration system, i.e. an arbitrary number of signs of the $\binom{N}{2}$ interactions of $\{J_{ij'}\}$ and $\{K_i^{\alpha'}\}$ in the no-frustration system are inverted. The subsets of $\{J_{ij'}\}$ and $\{K_i^{\alpha'}\}$, the signs of which are inverted, are randomly selected. As a result, frustrations are produced by degrees in the no-frustration system.

2.4. Exponentially attenuated interaction system

In this system, the interactions J_{ij} and K_i^{α} have the following dependence on the separation, $|i-j|$, along a chain

$$J_{ij} = bJ_{ij'}\,\mathrm{e}^{-A|i-j|}$$

$$K_i^{\alpha} = bK_i^{\alpha'}\mathrm{e}^{-B|i-j|} \qquad (3)$$

That is, the magnitude of an interaction is exponentially attenuated as the distance along a chain between the interacting two sites is increased. Here, $J_{ij'}$ and $K_i^{\alpha'}$ are homogeneous random numbers sampled within the interval $\{-1,1\}$, and A and B are arbitrary positive coefficients. If we take very large values for A and B, then this model is essentially the same as the one-dimensional Ising model.

Next, we consider the general behavior of the systems. Let an order parameter of a system be θ, and the free energy be a function of θ, $F(\theta)$. The value of θ is defined by conformations in the case of a polymer, e.g. θ measures the degree of formation of the native conformation of a polymer. Then

$$F(\theta) = E(\theta) - TS(\theta) \qquad (4)$$

where $E(\theta)$ and $S(\theta)$ are the energy and entropy of the system considered. In the case of a protein, we define θ_n as the value of the order parameter of the native state and θ_u as the value for the unfolding

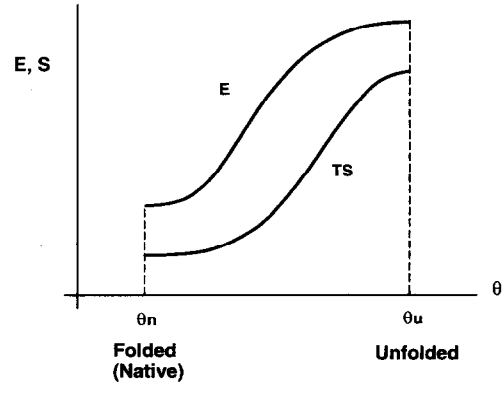

Fig. 1. Hypothetical dependence of E and S on an order parameter θ.

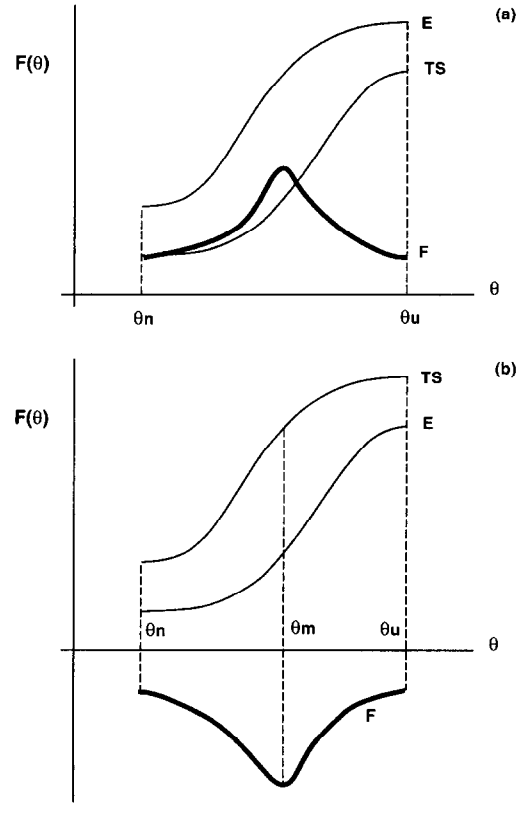

Fig. 2. (a) Schematic drawing of the free energy profile $F(\theta)$ (shown by a thick line) when the inflection point of the E curve (thin line) is on the left-hand side of the S curve (thin line). The $F(\theta)$ curve should have a barrier. (b) Schematic drawing of the free energy profile $F(\theta)$ (shown by a thick line) when the inflection point of the E curve (thin line) is on the right-hand side of the S curve (thin line).

state. It is natural to assume that E and S increase as θ deviates from θ_n and that E and S take minimum values at θ_n and maximum values at θ_u, the overall dependence of E and S on θ is approximately sigmoidal, as depicted in Fig. 1. When the inflection point of the S curve is on the right-hand side of that of the E curve, the $F(\theta)$ curve has a barrier between θ_n and θ_u, as shown in Fig. 2(a). Conversely, when the inflection point of the E curve is on the right-hand side of that of the S curve, the states θ_n and θ_u are unstable with regard to free energy, and the system tends to be an intermediate state θ_m, as presented in Fig. 2(b). Thus, when a system shows the two-state folding–unfolding transition, the $F(\theta)$ curve should have a barrier; in other words, the inflection point of the E curve should be on the left-hand side of that of the S curve.

We mention the relationship of our system to the convexity of the S–E curve proposed by Go [35] as a condition of the two-state transition. Let $\theta|_{Sr}$ and $\theta|_{Er}$ be the inflection points of the S and E curves, respectively. Then, at $\theta|_{Sr}$ and $\theta|_{Er}$

$$\left(\partial^2 E/\partial\theta^2\right)_{\theta|Er} = 0$$

and

$$\left(\partial^2 S/\partial\theta^2\right)_{\theta|Sr} = 0 \tag{5}$$

If $\theta|_{Sr} > \theta|_{Er}$ (Fig. 1), then $\partial^2 E/\partial\theta^2 < 0$ and $\partial^2 S/\partial\theta^2 > 0$ within the interval $\{\theta|_{Er}, \theta|_{Sr},\}$. We note then the following relationships

$$(\partial S/\partial E) = \frac{(\partial S/\partial\theta)}{(\partial E/\partial\theta)}$$

$$\left(\partial^2 S/\partial E^2\right) = \frac{1}{(\partial E/\partial\theta)^3}\left[\left(\partial^2 S/\partial\theta^2\right)(\partial E/\partial\theta) - (\partial S/\partial\theta)\left(\partial^2 E/\partial\theta^2\right)\right] \tag{6}$$

In Fig. 1, $(\partial E/\partial\theta) > 0$ and $(\partial S/\partial\theta) > 0$ always. Therefore

$$\left(\partial^2 S/\partial\theta^2\right)(\partial E/\partial\theta) > 0$$

$$-(\partial S/\partial\theta)\left(\partial^2 E/\partial\theta^2\right) > 0 \tag{7}$$

in this interval. Thus, $(\partial^2 S/\partial E^2) > 0$. This denotes the convexity of the S–E curve. Hence, the case of $\theta|_{Sr} > \theta|_{Er}$ is consistent with Go's model within the interval $\{\theta|_{Er}, \theta|_{Sr}\}$.

3. Procedure for the numerical calculation

3.1. Order parameter

The order parameter of an arbitrary spin configuration state is defined by the following equation when the spin set of the ground state is denoted as $\{\sigma_i^g\}$

$$\theta = \frac{1}{2N}\sum_{i=1}^{N}|\sigma_i - \sigma_i^g| \tag{8}$$

The value of θ of the spin configuration of a completely disordered state should be fairly close to 0.5. If a system exhibits a perfect two-state transition, the

system would take either a value of θ of 0.0 or of 0.5. In this case, an ensemble average value of θ, i.e. θ_a, should be related to the population ratio of the ground state in the ensemble. When this population ratio is denoted by C_g, the following relationships should hold

$$\theta_a = 0 \times C_g + (1 - C_g) \times 0.5$$

$$1 - 2\theta_a = C_g \tag{9}$$

The equilibrium constant K between the ground and the disordered states is then expressed as

$$K = \frac{1 - C_g}{C_g} \tag{10}$$

The van't Hoff plot, i.e. the $\ln K$ vs. $1/T$ plot of this system, should be a straight line with a slope proportional to ΔE if it follows a perfect two-state transition, namely, the transition behaves like a normal molecular reaction. In other words, the deviation of a van't Hoff plot of a system from a straight line denotes the degree of deviation of the behavior of the system from the two-state approximation.

3.2. Monte Carlo calculation

The Metropolis Monte Carlo (MC) method [36] was employed for numerical calculations. In the present MC calculation, each of the N spins was inverted one by one, and the energy of the system was calculated according to Eq. (1), followed by the Metropolis judgment in each time. We started the simulation at temperature $T + 3$. The previously described procedure was then iterated 500 times while decreasing the temperature to T exponentially for equilibration. An additional 1000 MC iterations were carried out at T, and we finally obtained a result of a simulation at T. We performed 50 runs of the whole procedure and took average values of several properties of the system.

To identify the spin configuration of the ground state and to evaluate its energy, the following calculations were also carried out. The same MC procedure was iterated 500 times while decreasing T from 3 to 0.01. Twenty runs using this procedure were performed, and the lowest energy state among these 20 states was regarded as the ground state.

4. Results

In the present study, we consider a spin system with 20 particles. As described in Section 2, the values of $\eta_{ij'}$, $J_{ij'}$ and $K_i^{\alpha'}$ were chosen as random numbers. The values of a and b were determined so that the ground-state energy falls roughly between -40 and -50.

4.1. Random system

We set parameters $a = 1.4$ and $b = 1.0$. The set of spins of the ground state of this system was identified as follows

$$\{\sigma_i\} = (-1,1,1,1,1,1,-1,-1,1,-1,1,1,1,$$
$$-1,1,-1,-1,1,-1,-1)$$

with energy -53.0.

The temperature dependence of the heat capacity

$$C = \frac{\langle E^2 \rangle - \langle E \rangle^2}{T^2} \tag{11}$$

is shown in Fig. 3(a). In the above equation, $\langle A \rangle$ denotes an average value of A. The dependence of E on the average order parameter, θ_a, is presented in Fig. 3(b) and the van't Hoff plot in Fig. 3(c). The plot in Fig. 3(c) shows a relatively sharp straight line with a certain slope at $T < 3.5$ $(1/T > 0.29)$, then changing discretely to show almost constant values $(\ln K \approx -0.5)$ at $T > 3.5$ $(1/T > 0.29)$. The implication is that this system tends to freeze into various states whose average value of $\ln K$ is -0.5, not a unique state, in this temperature region. Thus, this system does not have a stable ground state. In the MC simulation at $T = 0.5$, only 9 out of the 50 runs fell into the ground state. These facts suggest the existence of many states near the ground state.

The E–θ_a plot shown in Fig. 3(b) can be fitted by the following formula in the range $0.2 \leq \theta_a \leq 0.4$.

$$E = 87.7 - 1940.1\theta_a + 9794.8\theta_a^2 - 2.02 \times 10^4\theta_a^3$$
$$+ 1.48 \times 10^4\theta_a^4 \tag{12}$$

The inflection point of Eq. (12), $\theta_a|_{Er}$, is determined to be 0.264.

Here, we use the following relationship

$$S = \int \frac{C}{T}\,dT \tag{13}$$

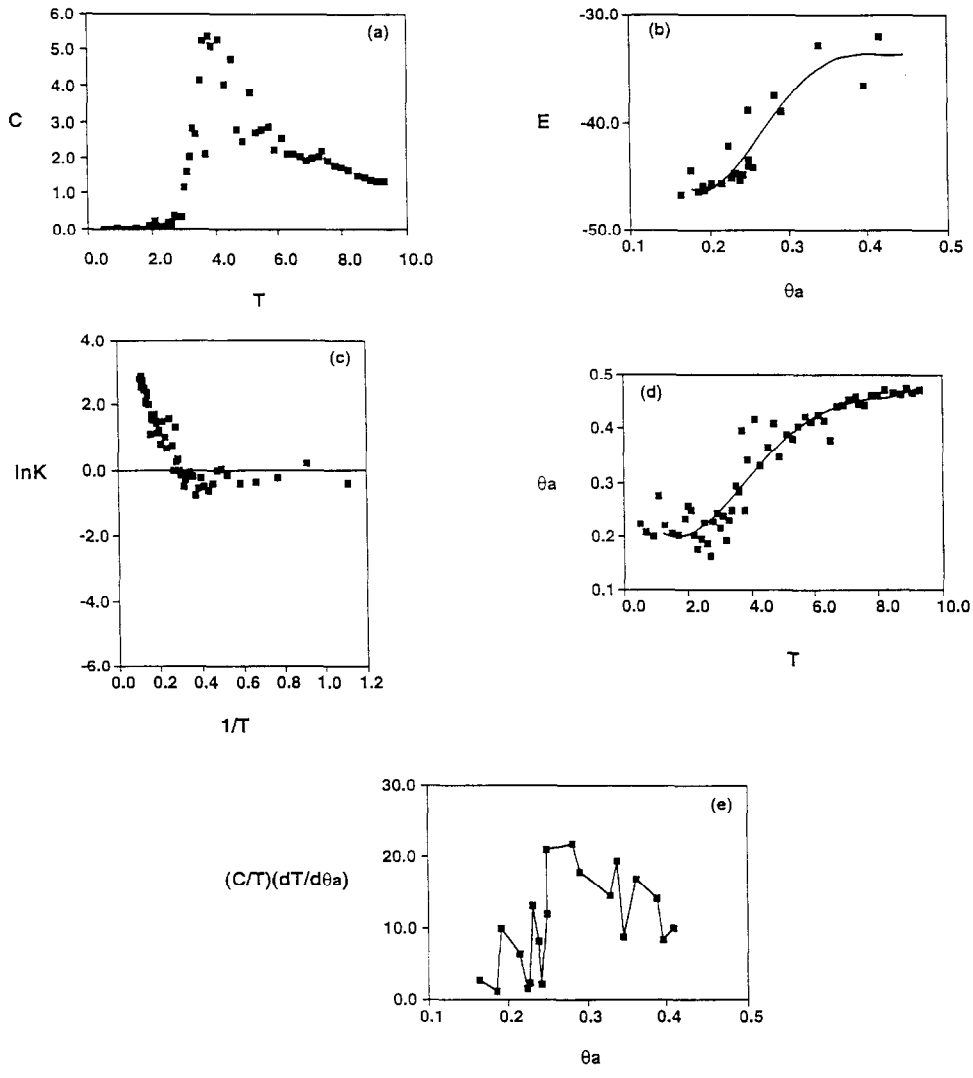

Fig. 3. Properties of the random system. (a) Temperature (T) dependence of C calculated using Eq. (11). (b) The θ_a dependence of E. The solid line denotes the curve fitted by Eq. (12). (c) Van't Hoff plot. (d). Temperature (T) dependence of θ_a. The solid line denotes the curve fitted by Eq. (16). (e). Plot of $(C/T)(\mathrm{d}T/\mathrm{d}\theta_a)$ vs. θ_a.

Eq. (13) is rewritten formally as

$$S = \int \frac{C}{T}\frac{\mathrm{d}T}{\mathrm{d}\theta_a}\,\mathrm{d}\theta_a \tag{14}$$

Therefore, at $\theta_a|_{Sr}$

$$\frac{\partial^2 S}{\partial \theta_a^2} = \frac{\partial}{\partial \theta_a}\left(\frac{C}{T}\frac{\mathrm{d}T}{\mathrm{d}\theta_a}\right) = 0 \tag{15}$$

Namely, the inflection point of S can be given by the peak of the curve defined by $(C/T)(\mathrm{d}T/\mathrm{d}\theta_a)$.

We can also approximate the θ_a–T plot shown in Fig. 3(d) using Eq. (16) in the range $2.0 \le T \le 8.0$.

$$\theta_a = 0.319 - 0.174T + 00761T^2 - 0.00992T^3 + 0.000424T^4 \tag{16}$$

The corresponding range of θ_a is [0.2,0.44]. Thus, the profile of $(C/T)(\mathrm{d}T/\mathrm{d}\theta_a)$ can be obtained from Eq. (16), as shown in Fig. 3(e). The highest peak in this plot is located at $\theta_a = 0.287$. This point corresponds to the inflection point of the S curve, $\theta_a|_{Sr}$.

The positions of $\theta_a|_{Er}$ and $\theta_a|_{Sr}$ are not much different. Based on the discussion in Section 2, this result suggests that a clear two-state transition in this system cannot be expected, and the behavior of the plot in Fig. 3(c) attests to this expectation.

Fig. 4 shows the change in profile of the population of spin states by temperature. We classified the population of final spin configuration states (denoted by θ) obtained in the 50 simulations. The profile of the population is essentially the same between $T=0.5$ and $T=3.5$, as seen in Fig. 4(a) and 4(b). The states are distributed in several spin configuration states in this temperature range. In other words, each simulation freezes into one of several glassy states below the transition temperature ($T \approx 3.5$). It is observed that the states are further randomized broadly at $T > 5.0$, higher than the transition temperature.

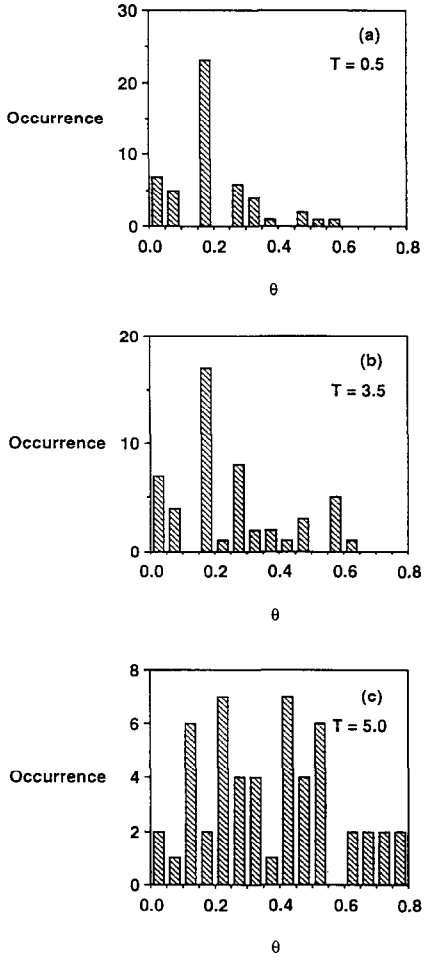

Fig. 4. The temperature dependence of the population of spin configuration states (θ) of the random system. (a) $T=0.5$. (b) $T=3.5$ (c) $T=5.0$.

4.2. No-frustration system

The values of parameters for this system were taken as $a=0.56$ and $b=0.4$. The ground state energy of the system was -48.1 with the following spin set

$$\{\sigma_i\} = (1,1,-1,-1,1,1,1,-1,-1,-1,1,1,1,1,1,-1,-1,1,-1,-1)$$

This ground state is quite stable, and all 50 runs fell into this state in the simulation at $T \leq 0.8$. The properties of this system corresponding to Fig. 3(a)–3(e) in the case of the random system are presented in Fig. 5(a)–5(e). The $E-\theta_a$ profile (Fig. 5(b)) can be approximated by Eq. (17) in the region $0 \leq \theta_a \leq 0.3$.

$$E = -49.57 + 125.53\theta_a + 436.92\theta_a^2 - 3701.3\theta_a^3 + 6257.4\theta_a^4 \quad (17)$$

The inflection point of Eq. (17), $\theta_a|_{Er}$, is located at 0.0466. The θ_a-T plot in Fig. 5(d) can be fitted by the following formula within $1.5 \leq T \leq 6.0$

$$\theta_a = 0.535 - 0.698T + 0.305T^2 - 0.0486T^3 + 0.00280T^4 \quad (18)$$

The location of the highest peak in $(C/T)(dT/d\theta_a)$ calculated from Eq. (18) is at $\theta_a = 0.162$, as shown in Fig. 5(e). That is, the inflection point $\theta_a|_{Sr}$ is clearly on the right-hand side of $\theta_a|_{Er}$. Thus, this system is expected to show a clear two-state transition. Corresponding to this analysis, the van't Hoff plot presented in Fig. 5(c) exhibits remarkable linearity, especially at $0.2 \leq 1/T \leq 0.6$ ($1.67 \leq T \leq 5.0$).

We would like to comment on the heat capacity of the system at low temperature. As previously mentioned, the random system fell into one of several states with various values of the heat capacity below the transition temperature. In the no-frustrated system, all 50 runs freeze into the lowest energy state below $T=0.5$ showing a 0 heat capacity value after the equilibration process described in Section 3

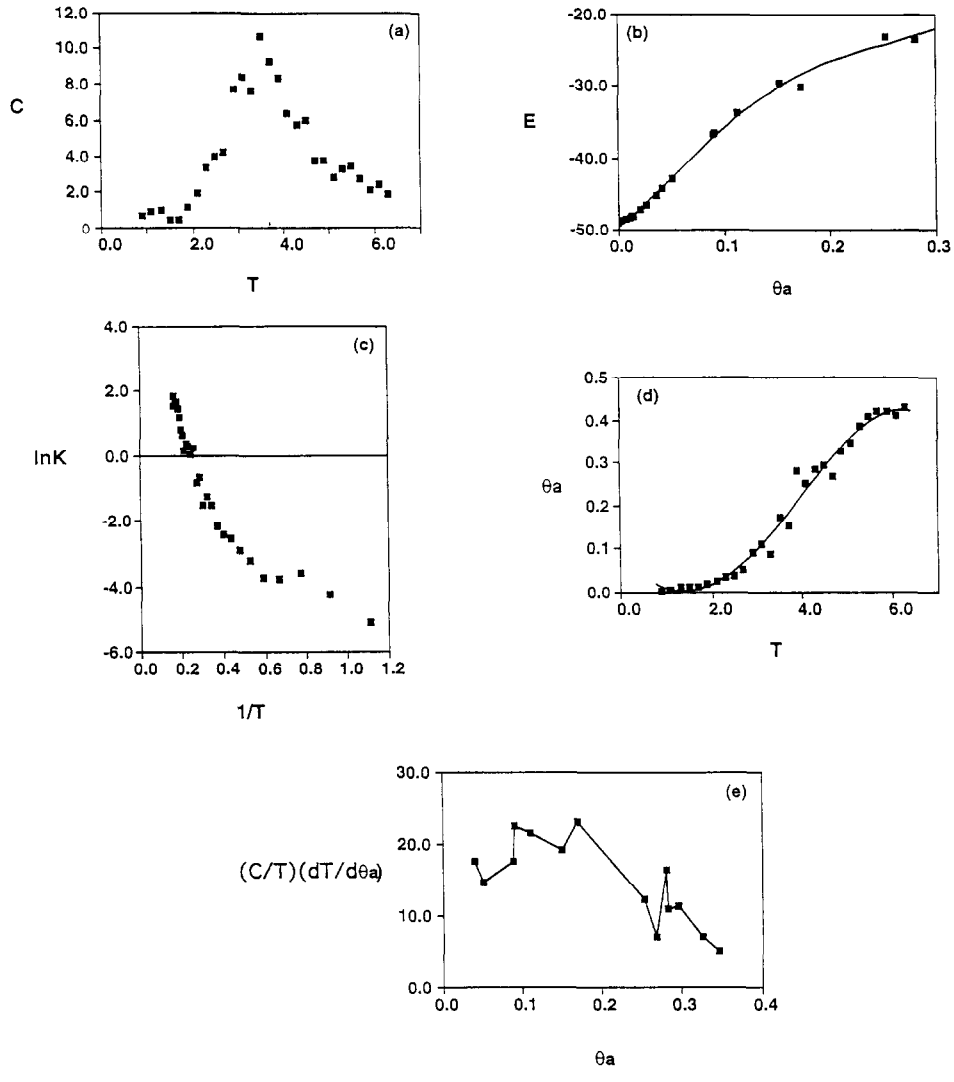

Fig. 5. Properties of the no-frustration system. (a) Temperature (T) dependence of C calculated using Eq. (11). (b) The θ_a dependence of E. The solid line denotes the curve fitted by Eq. (17). (c) Van't Hoff plot. (d). Temperature (T) dependence of θ_a. The solid line denotes the fitted curve by Eq. (18). (e). $(C/T)(dT/d\theta_a)$ vs. θ_a plot.

(i.e. no thermal fluctuation). At $T = 1.0$, 17 of the 50 runs fall into the same state. The other 33 runs show thermal fluctuation with the heat capacity value 0.940. This demonstrates that the 33 simulations reached the same thermal fluctuation state.

The relationship between the population of spin states and the temperature is presented in Fig. 6. At $T = 0.5$, all states are in the lowest energy state (i.e. $\theta = 0.0$). As T is increased, denatured states are gradually accumulated, while the lowest energy state is still highly populated. At $T = 3.5$, i.e. around the transition temperature estimated from the heat capacity plot in Fig. 5(a), the number of states with $0.0 \leq \theta_a \leq 0.05$ is 26 out of 50, and other states are broadly distributed with a peak between 0.1 and 0.15. At $T = 5.0$, the system is completely randomized, as shown in Fig. 6(c). This property also indicates the two-state nature of the transition.

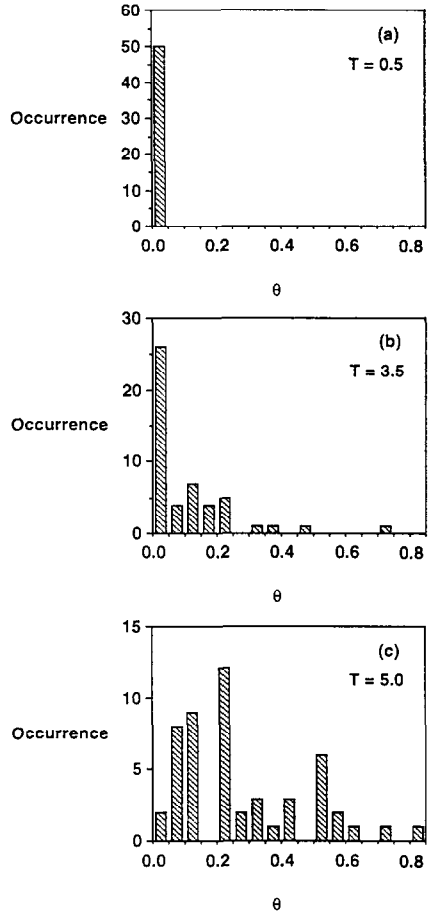

Fig. 6. The temperature dependence of the population of spin configuration states (θ) of the no-frustration system. (a) $T = 0.5$. (b) $T = 3.5$. (c) $T = 5.0$.

4.3. Partially frustrated systems

According to the method described in Section 2.3, we introduce frustrations gradually into the no-frustration system previously defined. We inverted the signs of m arbitrary interactions of $\{J_{ij}\}$ and $\{K_i^{\alpha}\}$. We tried the cases of $m = 5$, 10, 15, 20, and 25. Fig. 7(a) shows the van't Hoff plots of the frustrated systems, the random system, and the no-frustration system. In general, as the number of frustrations increases, the van't Hoff plot deviates from linearity and comes close to the profile of the random system. The same spin configuration of the ground state of the no-frustration system is also the lowest energy state in each case for $m = 5$–20. In the system with $m = 25$, one spin in the lowest state is inverted compared with the ground state of the no-frustration system, but these two states have similar energy values (-39.8 and -38.3, respectively).

Fig. 7(b) summarizes the location of the $\theta_a|_{Er}$ and $\theta_a|_{Sr}$ in each case. The distance between $\theta_a|_{Er}$ and $\theta_a|_{Sr}$ tends to be smaller as m increases, in general. That is, the partially frustrated systems come close to the random system for increasing m.

These results suggest that the lowest energy state becomes unstable, gradually keeping its two-state nature by the stepwise introduction of frustrations, and finally the unique lowest energy state disappears and many alternative low-energy states appear. We think that this property of our systems corresponds to the mutational properties of proteins discussed by Lattman and Rose [23] and Nishikawa [24].

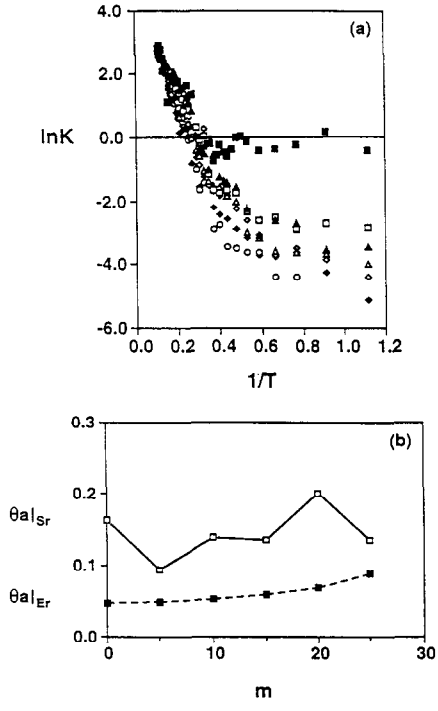

Fig. 7. (a) Van't Hoff plot of (■) the random system, (♦) the no-frustration system, and the partially frustrated systems: (○) $m = 5$, (◇) $m = 10$, (△) $m = 15$, (F□) $m = 20$, and (▲) $m = 25$. (b) Plots of (□) $\theta_a|_{Sr}$ and (■) $\theta_a|_{Er}$ vs. the number of frustrations (m).

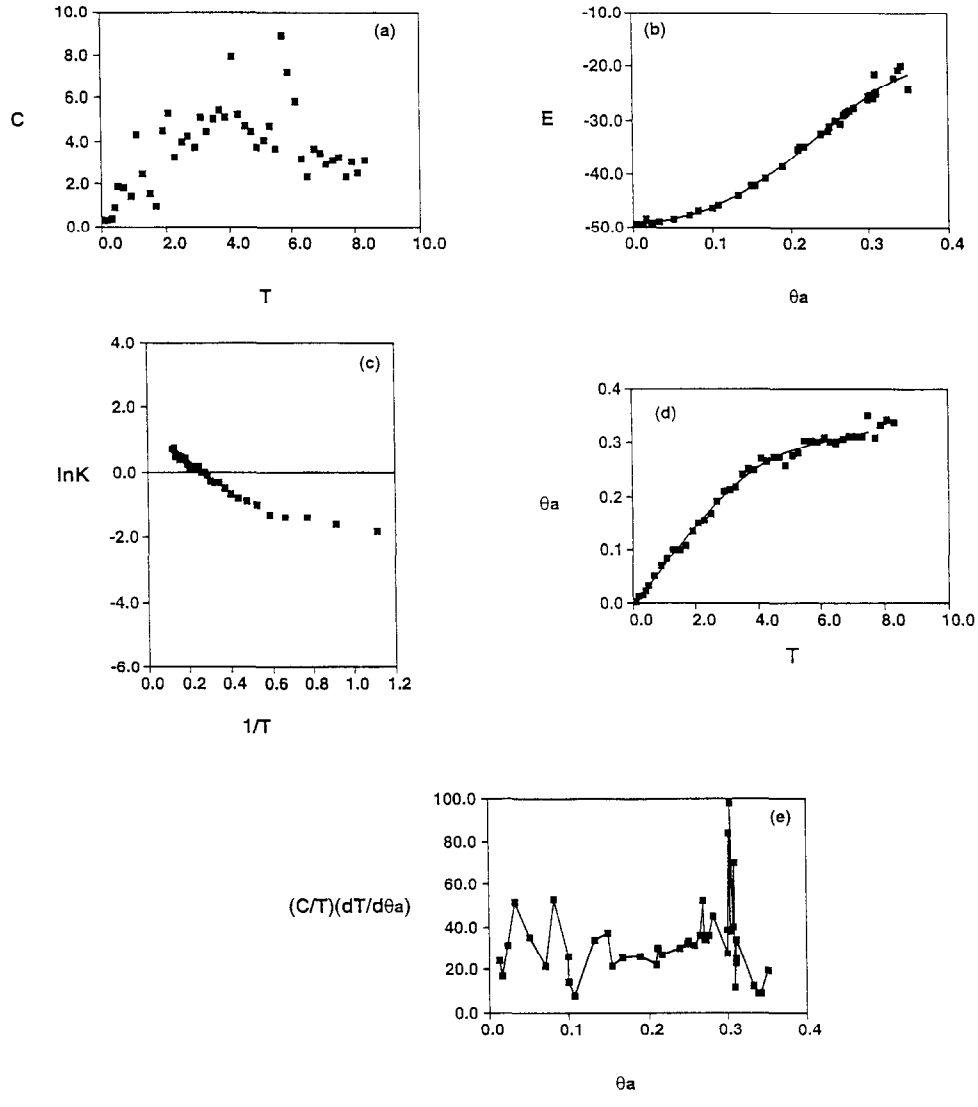

Fig. 8. Properties of the exponentially attenuated interaction system. (a) Temperature (T) dependence of C calculated using Eq. (11). (b) The θ_a dependence of E. The solid line denotes the curve fitted by Eq. (19). (c) Van't Hoff plot. (d). Temperature (T) dependence of θ_a. The solid line denotes the curve fitted by Eq. (20). (e). $(C/T)(dT/d\theta_a)$ vs. θ_a plot.

4.4. Exponentially attenuated interaction system

We took the parameters in Eq. (4) as $a = 5.0$, $b = 2.5$ and $A = B = 2.0$. The energy of the ground state of this system was -49.5. The following spin set was obtained as the ground state

$$\{\sigma_i\} = (1,1,-1,-1,1,1,1,1,-1,-1,-1,1,1,1,1,1,1,-1,-1,1,-1,-1)$$

All states in the simulations at $T = 0.01$ fell into the ground state. In this sense, this system has a unique stable state. However, in the simulations at $T = 0.1$, only 20 states out of 50 runs converged to the ground state. It is suggested that this ground state is stable but easily disturbed by heating.

Fig. 8(a)–8(e) represent the properties of this system. Both the E–θ_a and θ_a–T plots in Fig. 8(b) and 8(d) seem to be rather smooth. The E–θ_a plot

can be fitted within the interval $0.05 \leq \theta_a \leq 0.35$ if the following formula is used for the approximation

$$E = -49.29 + 6.251\theta_a + 167.6\theta_a^2 + 1057.6\theta_a^3 - 2694.2\theta_a^4 \quad (19)$$

The inflection point of Eq. (19), $\theta_a|_{Er}$, is 0.24. We can also approximate the plot of θ_a–T in Fig. 8(d) within the interval $0.1 \leq T \leq 7.0$ by Eq. (20)

$$\theta_a = 0.0174 + 0.0668T + 0.00784T^2 - 0.00296T^3 - 0.000199T^4 \quad (20)$$

This temperature range corresponds to the interval $0.001 \leq \theta_a \leq 0.3$.

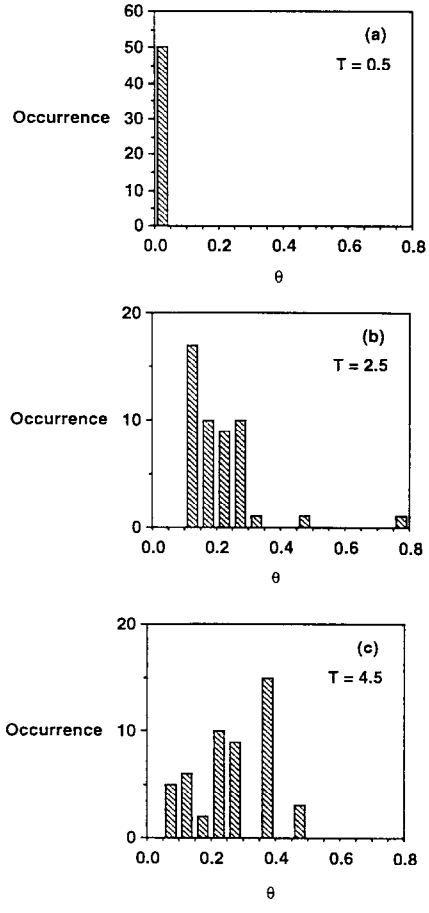

Fig. 9. The temperature dependence of the population of spin configuration states (θ) of the exponentially attenuated interaction system. (a) $T = 0.5$. (b) $T = 2.5$. (c) $T = 4.5$.

The T dependence of C (Fig. 8(a)) is fairly complicated compared with the random and no-frustration systems. There are several distinct peaks. The corresponding $(C/T)(dT/d\theta_a)$ vs. θ_a plot shown in Fig. 8(e) also has several peaks. However, two distinct peaks can be recognized at 0.034 and 0.083. These peaks are both clearly on the left-hand side of $\theta_a|_{Er} = 0.24$. This situation corresponds to the case described with Fig. 2(b) in Section 2. That is, it is suggested that this system shows a smooth transition at least in the range $0 \leq \theta_a \leq 0.24$. Actually, the linearity of the van't Hoff plot of this system is not so remarkable when compared with the no-frustration system, as seen in Fig. 8(c).

The temperature dependence of the population of the spin states is shown in Fig. 9. All simulations converge to states with $0.0 \leq \theta \leq 0.05$ at $T = 0.5$. As the temperature is increased from $T = 0.5$ to $T = 4.5$, the peak of the histogram of the distribution shifts continuously from $\theta \approx 0.12$ to $\theta \approx 0.37$, as presented in Fig. 9(a)–9(c). The location of the peak moves to higher θ values as the temperature increases. This behavior indicates the characteristics of a loose transition.

5. Discussion

As observed in the comparison between the no-frustration and the random systems in Section 4, when $\theta_a|_{Er} < \theta_a|_{Sr}$, a two-state transition of the no-frustration system is clearly detectable. The condition for the two-state transition proposed in Section 2 is held in the present spin systems.

The no-frustration system shows the basic properties that a protein possesses. We confirmed the uniqueness of the ground state, namely, one of the characteristics of proteins, by simulations below $T = 0.8$ where each simulation converged to the ground-state spin configuration. The most important property is the two-state transition of the system. This behavior of the system was observed as the linearity of the van't Hoff plot, especially at $1/T < 0.5$ (see Fig. 5(c)). It should be noted that we do not treat the so-called cold denaturation of proteins in this work. We confirmed that a system with a different energy

parameter set without frustration exhibits the property of the two-state transition. That is, this property is attributed essentially to its no-frustration character.

It was also observed in our simulations that these properties are preserved in the partially frustrated systems to a certain degree, but the systems are generally close to the random system, depending on the degree of frustration. It seems that this behavior of frustrated systems corresponds to that of mutated proteins. As Lattman and Rose [23] indicated, the extremely low energy of the native structure compared with other states is not the only factor which determines the native conformation of a protein. On the contrary, a certain degree of destabilization of the native structure by mutations does not significantly affect the two-state transition between the native and the unfolded states. They also pointed out that the structure of the native state is never changed to another completely different stable structure, for example from lysozyme to RNase A. The sufficiently low energy of the native state from the first excited state is not enough to explain this property of proteins, although this condition is one of the important factors for protein folding [16,17].

The two-state behavior and mutational properties of real proteins may be interpreted based on our model as follows. Assume that there is a protein sequence whose tertiary structure has no frustration. The native structure of this sequence would be quite stable. The relationship between the inflection points, $\theta_a|_{Er} < \theta_a|_{Sr}$, should hold in this ideal state. The stability of this native structure would decrease by mutations with the addition of frustrations. Finally, the native state and states conformationally near the native state become energetically almost degenerate after the introduction of sufficient numbers of mutations. That is, the addition of frustrations decreases the relative distance between $\theta_a|_{Er}$ and $\theta_a|_{Sr}$, and finally the positions of these two inflection points overlap. The protein finally cannot fold into a unique structure but behaves as a random polypeptide. In this sense, the no-frustration state can be regarded as an ideal state of a protein, as pointed out by Go [6] and Bryngelson and Wolynes [26]. In this connection, it should be noted that many different sequences of proteins fold into essentially the same structure. These sequences would also satisfy the condition $\theta_a|_{Er} < \theta_a|_{Sr}$.

An exponentially attenuating interaction system also possesses a stable ground state, but this is easily disordered by an increase in the temperature. Both the $E-\theta_a$ plot (Fig. 8(b)) and the thermal transition profile (Fig. 8(d)) show relatively loose curves. Thus, the two-state transition is not distinct in this system. Several peaks in the $C-T$ plot or the $(C/T)(dT/d\theta_a)-\theta_a$ plot (Fig. 8(a) and 8(e)) suggest that the melting process contains several transitions. These properties are similar to those of nucleic acids [4], i.e. this system also provides a simple view of the behavior of nucleic acids in terms of the transition between the unique native structure and the denatured structure. As seen in the $(C/T)(dT/d\theta_a)-\theta_a$ plot in Section 4.4, the two distinct peaks are clearly on the left of the position of $\theta_a|_{Er}$, i.e. $\theta_a|_{Sr} < \theta_a|_{Er}$. Although we think that the loose transition of this system reflects this fact, in other words, the condition of a loose transition proposed in Section 2 is satisfied by this system, the situation seems to be more complicated than the model described with Fig. 2(b) in Section 2. The two-state transition requires long-range interactions between two sites, as many authors have pointed out [9,18,37]. This statement is consistent with the present results of the no-frustrated system and the exponentially attenuated interaction system. Thus, the present spin systems, which are expressed by the same Hamiltonian, Eq. (1), show the properties both of proteins and of nucleic acids. That is, the difference in the interactions between two sites accounts for the difference between the properties of proteins and nucleic acids. We consider that the system described by Eq. (1) offers a unified view of biopolymers.

To perform a calculation such as those presented in this paper for real proteins, we must explicitly define an order parameter. The most plausible candidates for an order parameter seem to be the radius of gyration or the overlap function used in the replica approach of protein folding by Shakhnovich and Gutin [29]. However, in a practical sense, further development of a technique for efficient conformational sampling on a potential surface is required in order to complete the calculation for a real protein. Although it seems to be very difficult (there have been some new attempts to solve this problem — see, for example, Refs. [21] and [40]), we are planning to apply the present analysis to small proteins.

References

[1] C.B. Anfinsen and H.A. Scheraga, Adv. Protein Chem., 29 (1975) 205.
[2] G. Némethy and H.A. Scheraga, Q. Rev. Biophys., 10 (1977) 239.
[3] P.L. Privalov and S.J. Gill, Adv. Protein Chem., 39 (1988) 191.
[4] Y. Maeda, K. Takahashi, H. Yamaki and E. Ohtsubo, Biopolymers, 27 (1988) 1917.
[5] A. Rich and U.L. RajBahansary, Annu. Rev. Biochem., 45 (1976) 805.
[6] N. Go, Annu. Rev. Biophys. Bioeng., 12 (1983) 183.
[7] M. Karplus and E.I. Shakhnovich, Protein Folding, Freeman, New York, 1992.
[8] H. Taketomi, F. Kano and N. Go, Biopolymers, 27 (1988) 527.
[9] J. Skolnick and A. Kolinski, Annu. Rev. Phys. Chem., 40 (1989) 207.
[10] J. Skolnick and A. Kolinski, J. Mol. Biol., 221 (1991) 499.
[11] H.S. Chan and K.A. Dill, J. Chem. Phys., 95 (1991) 3775.
[12] H.S. Chan and K.A. Dill, J. Chem. Phys., 100 (1994) 9238.
[13] D.G. Covel and R.L. Jernigan, Biochemistry, 29 (1990) 3287.
[14] D.G. Covel, Proteins, 14 (1992) 409.
[15] D.G. Covel, J. Mol. Biol., 235 (1994) 1032.
[16] A. Sali, E.I. Shakhnovich and M. Karplus, Nature, 369 (1994) 248.
[17] A. Sali, E.I. Shakhnovich and M. Karplus, J. Mol. Biol., 235 (1994) 1614.
[18] H. Taketomi, Y. Ueda and N. Go, Int. J. Peptide Protein Res., 7 (1975) 445.
[19] E.I. Shakhnovich and A.V. Finkelstein, Biopolymers, 28 (1989) 1667.
[20] A.V. Finkelstein and E.I. Shakhnovich, Biopolymers, 28 (1989) 1681.
[21] M.-H. Hao and H.A. Scheraga, J. Phys. Chem., 98 (1994) 4940.
[22] M.-H. Hao and H.A. Scheraga, J. Phys. Chem., 98 (1994) 9882.
[23] E.E. Lattman and G.D. Rose, Proc. Natl. Acad. Sci. USA, 90 (1993) 439.
[24] K. Nishikawa, Viva Origino, 21 (1993) 91.
[25] D.L. Stein, Proc. Natl. Acad. Sci. USA, 82 (1985) 3670.
[26] J.D. Bryngelson and P.G. Wolynes, Proc. Natl. Acad. Sci. USA, 84 (1987) 7524.
[27] D. Bryngelson and P.G. Wolynes, J. Phys. Chem., 93 (1989) 6902.
[28] E.I. Shakhnovich and A.M. Gutin, Biophys. Chem., 34 (1989) 187.
[29] E.I. Shakhnovich and A.M. Gutin, Nature, 346 (1990) 773.
[30] V.I. Abkevich, A.M. Gutin and E.I. Shakhnovich, J. Chem. Phys., 101 (1994) 6052.
[31] R.M. Wartell and E.W. Montroll, Adv. Chem. Phys., 22 (1972) 129.
[32] N.S. Gocl and E.W. Montroll, Biopolymers, 6 (1968) 731.
[33] B.H. Zimm and J.K. Bragg, J. Chem. Phys., 31 (1959) 526.
[34] D. Poland and H.A. Scheraga, Theory of Helix Coil Transition in Biopolymers, Academic Press, New York, 1970.
[35] N. Go, Int. J. Peptide Protein Res., 7 (1975) 313.
[36] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller and E. Teller, J. Chem. Phys., 21 (1953) 1087.
[37] V.I. Abkevich, A.M. Gutin and E.I. Shakhnovich, J. Mol. Biol., 252 (1995) 460.
[38] J.N. Onuchic, P.G. Wolynes, Z. Luthey-Schulten and N. Socci, Proc. Natl. Acad. Sci., 92 (1995) 3626.
[39] J.D. Bryngelson, J.N. Onuchic, N.D. Socci and P.G. Wolynes, Proteins, 21 (1995) 167.
[40] U. Hansmann and Y. Okamoto, J. Comput. Chem., 14 (1993) 1333. E.M. Boczko and C.L. Brooks, III, Science, 269 (1995) 393.