ResearchGate

See discussions, stats, and author profiles for this publication at:
https://www.researchgate.net/publication/19786627

The effect of amino acid substitution on protein-folding and -unfolding transition studied by computer simulation

ARTICLE *in* BIOPOLYMERS · MAY 1988

Impact Factor: 2.39 · DOI: 10.1002/bip.360270402 · Source: PubMed

CITATIONS
35

READS
8

3 AUTHORS, INCLUDING:

Fumiaki Kanô
Showa University
17 PUBLICATIONS **162** CITATIONS

SEE PROFILE

Nobuhiro Go
Japan Atomic Energy Agency
238 PUBLICATIONS **9,848** CITATIONS

SEE PROFILE

Available from: Nobuhiro Go
Retrieved on: 15 January 2016

The Effect of Amino Acid Substitution on Protein-Folding and -Unfolding Transition Studied by Computer Simulation

HIROSHI TAKETOMI, *Computer Center, Kyushu University 91, Fukuoka 812, Japan,* FUMIAKI KANÔ, *Department of Physics, College of Arts and Sciences, Showa University, Fuji-Yoshida 403, Japan,* and NOBUHIRO GŌ, *Department of Chemistry, Faculty of Science, Kyoto University, Kyoto 606, Japan*

Synopsis

Protein-folding and -unfolding transitions were studied by the method of computer simulation. The protein was modeled as a two-dimensional lattice polymer. Various energy terms were assumed to be operative between units composing the polymer. But hydrophobic interactions were neglected explicitly. Both thermodynamic and kinetic quantities were obtained from the simulation, and from their temperature dependence in the transition zone characteristics of the conformational transition of proteins were discussed. Two amino acid substituted models, differing in the location of substitution, were studied and compared with the original in order to clarify the effect of substitution on conformational transition of proteins. The following conclusions were reached in this study: (1) The relaxation time of the slow mode, which reflects the overall folding and unfolding processes, shows a peak near the transition temperature, while that of the fast mode is almost independent of temperature. The peak of the slow mode occurs at a slightly lower temperature than the transition temperature. (2) The dependence of the logarithm of the rate constants on the inverse of temperature (Arrhenius plot) is linear. Therefore, the plot of the free energy of activation vs temperature is linear. (3) The values of kinetic parameters obtained suggest that in the activated state the intramolecular interactions are half broken, while the state is close to the native state on the entropy axis. (4) The amino acid substitution, which is modeled as having slightly unfavorable short-range interactions, causes the substituted ones to be slightly unstable. Moreover, it causes the folding transition to slow. From the analysis of the way slowing down is observed in the two substituted models, we conclude that a structure, designed to model a β-sheet, is formed before it gets assembled with other structures, which are designed to model α-helices. The process of assembly occurs nearly at the activated state of the folding and unfolding transition. (5) It is suggested from this study that the maximum of folding rate constant in the Arrhenius plot that has been observed experimentally in real proteins is likely due to hydrophobic interactions.

INTRODUCTION

Elucidation of the mechanism of self-organization of protein native structure[1-4] is one of the ultimate aims of the study of protein conformation. We have been approaching this problem by introducing a rather abstract simplified theoretical model and by studying it from the statistical mechanical point of view, i.e., we introduce lattice models of proteins with various interaction terms that are essential to determine various properties of protein conformation.[5-10] As usually done in polymer physics, we divide the interactions conceptually into two categories, long-range interactions and short-range

Biopolymers, Vol. 27, 527–559 (1988)
© 1988 John Wiley & Sons, Inc.
CCC 0006-3525/88/040527-33$04.00

ones. Long-range interactions are the ones that are operative between units separated far along the chain, but usually near in space. Short-range interactions are those between units near along the chain. In addition, we take into consideration other important interactions—hydrophobic ones—in a simplified manner as described later. By using such a model and the interactions, we have been studying the respective roles of these three types of interactions in the folding and unfolding transition.[5-10]

In this paper, we treat a two-dimensional lattice model of protein and study it by the method of computer simulation. In this lattice model, the excluded-volume effect is taken into account by checking the overlap of the units. We especially focus our attention on temperature effects upon thermodynamics and kinetics of protein folding and unfolding in the transition zone. Using an approach with such a highly simplified model, we can study the prototype of the transition in a unified manner from both thermodynamic and kinetic aspects. We will consider three types of model proteins. The first one is called "Original." The other two are amino acid substituted models ("Substitution 1" and "Substitution 2"). The substituted models, which are modeled as having slightly unfavorable short-range interactions, are studied in order to clarify the effect of small perturbations in amino acid sequences on the folding properties of the protein.

Here we describe only the results obtained for the case in which hydrophobic interactions[11] are neglected. Even though we confine ourselves to the case of absence of the hydrophobic interactions in this paper, it is possible to discuss effects of the hydrophobic interactions on the folding and unfolding transition by comparing the results of this simulation with real experimental data. Because the hydrophobic interactions have an effect of decelerating the transition,[9] more computing time for longer simulation will be needed. Such a simulation is now in progress.

To be brief, the purposes of this study are as follows:

(1) To study the prototype of the equilibrium and kinetic characteristics of protein conformational transition.

(2) To elucidate the relations between the equilibrium/kinetic properties of the transition and the three types of interactions (long-range, short-range, and hydrophobic interactions).

(3) To clarify the effect of amino acid substitution on the equilibrium and kinetic behavior of the transition.

In the first section, we describe our theoretical framework. We introduce a lattice model of proteins and the method of Monte Carlo simulation. Then, the methods of analysis of the simulation data are described. In the second section, we present and discuss our results. The temperature dependence of relaxation times and rate constants (Arrhenius plot) is discussed, especially for the latter from the point of view of activated state of the transition. Comparisons are made of the temperature dependence of the rate constants obtained by two different studies, this simulation and the experimental ones in real proteins. The latter shows a peculiar convex dependence. Based on these results, the previous theories on these phenomena are discussed and our viewpoint is offered. For amino acid substituted models, the results are compared and the effects of substitutions are discussed. In the last section, conclusions of our study are given.

THEORETICAL FRAMEWORK

Lattice Model of Protein

A lattice protein molecule is a self-avoiding chain polymer on the two-dimensional square lattice[5,7-9] or on the three-dimensional cubic lattice,[6,10] but here we consider only the two-dimensional case. The native conformation that we adopt in this paper is shown in Fig. 1. It consists of 49 ($= 7 \times 7$) units linked linearly by bonds. It can assume a variety of conformations depending on a bond angle at each unit. However, by specific interactions introduced below, which we regard as originating from genetic information, this lattice protein assumes a unique specific conformation, that is, native conformation, at low temperatures. The uniqueness of the native conformation ensures that we can unambiguously judge whether the conformation reaches the final goal or not.

In a series of papers on the simulation of the lattice model of protein, we have been considering three types of intramolecular interactions, as mentioned in the introduction.

(a) Long-range interactions[5] are modeled as follows: Attractive interactions are assumed to be operative among a set of preassigned pairs of units, when they occupy nearest neighbor lattice points. Such pairs are called "interactable pairs." In the most specific case, the interactable pairs are chosen exactly the same as the ones in the native conformation in Fig. 1. In this paper we adopt this most specific case as long-range interactions. The distribution of interactable pairs in this case is shown in Fig. 2 and interactable ones are indicated by solid (black) squares. The number of the black squares in Fig. 2 is 36. The designation of the long-range interactions is specific in the sense that they selectively stabilize the native conformation shown in Fig. 1. When one

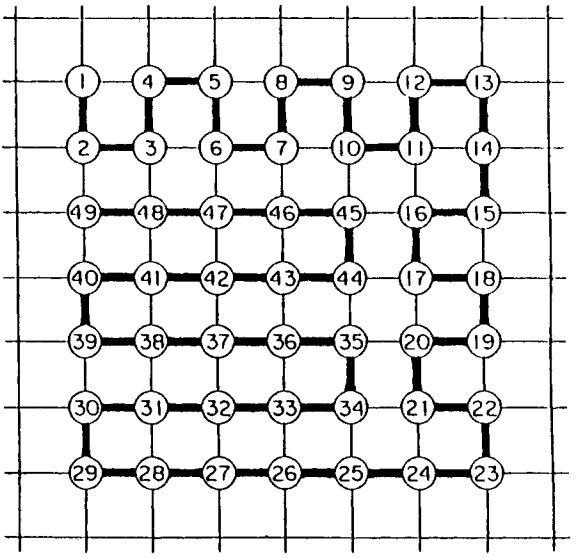

Fig. 1. The native conformation of the lattice protein on the two-dimensional square lattice.

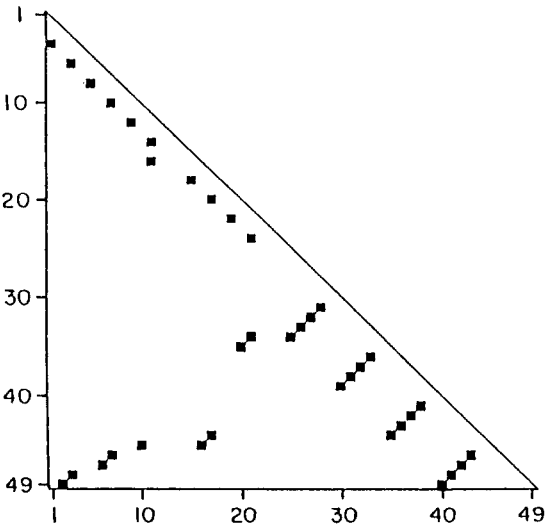

Fig. 2. The designation of long-range interactions in the lattice protein. Both abscissa and ordinate indicate unit number. Interactable paris of units are indicated by black squares.

pair of interactable units occupies nearest neighbor lattice points, we assume that the energy of the system decreases by ϵ_1.

The conformation in Fig. 1 has two-dimensional analogues of secondary structures, that is, two α-helices (from units 1 to 14, and units 15 to 24) and a β-sheet (from units 25 to 49). The regular structure is reflected on the map of the long-range interactions in Fig. 2. The pairs from units 1 to 24 along the diagonal line indicate the pattern of the α-helices. The ones from units 25 to 49 perpendicular to the diagonal line indicate that of the β-sheet. The other pairs are located far from the diagonal line. They stand for the interactions between the α-helices and the β-sheet.

(b) Short-range interactions[8] are modeled as the tendency for each unit to take the bond angle similar to that in the native conformation. Therefore, we assume that short-range interactions are operative and the energy of the system decreases by ϵ_s, when a local conformation consisting of three consecutive units has the same bond angle as in the native one in Fig. 1. The designation of the short-range interactions is shown in Fig. 3. The interactions work when each unit takes the same bond angle as the one indicated by solid square in Fig. 3. These interactions also contribute in specifically stabilizing the native conformation.

(c) Hydrophobic interactions[9] are modeled as follows: First we classify each unit into two categories, polar or nonpolar. We regard any vacant lattice sites in any conformation of lattice protein as occupied by water molecules. Therefore, we assume that hydrophobic interactions are operative when the sites next to nonpolar units are vacant and the entropy of the system decreases proportionally to the number of vacant sites next to nonpolar units. This causes indirect attractive interactions between nonpolar units because of

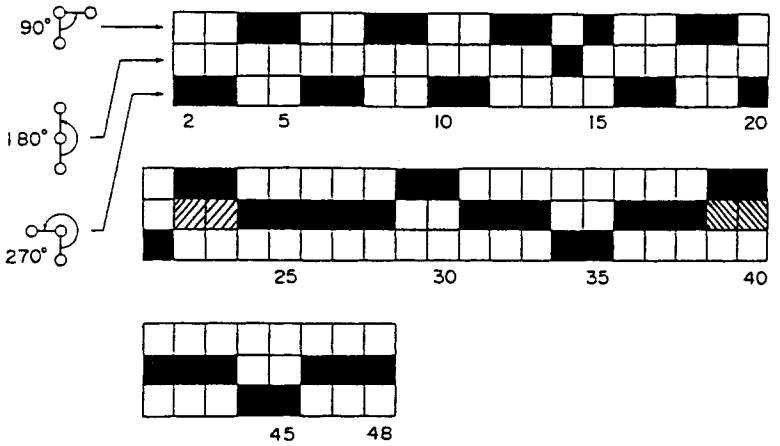

Fig. 3. The designation of short-range interactions in the lattice protein. Substitution 1 has substituted units at the 39- and 40th unit, where the bond angles indicated by hatched squares are more favorable ones. Substitution 2 has substituted units at the 22- and 23th unit with more favorable bond angles indicated by hatched squares.

the free energy changes of the system. These attractive interactions have a nonspecific nature in contrast to the long- or short-range interactions introduced above. It is well known that the hydrophobic interactions play an important role in determining the properties of protein, but in this paper we confine ourselves to the case where the hydrophobic interactions are explicitly neglected for the reasons discussed in the introduction.

Based on the interactions described above, the conformational energy (enthalpy) H of the lattice protein system is given by

$$H = -(m_l\epsilon_l + m_s\epsilon_s) \tag{1}$$

where m_l and m_s denote the number of occurrences of long- and short-range interactions, respectively. We introduce the quantities ϵ_0, w_l and w_s by the relations $\epsilon_l = w_l\epsilon_0, \epsilon_s = w_s\epsilon_0$, and $\epsilon_0 = \epsilon_l + \epsilon_s$. Then,

$$H = -(m_l w_l + m_s w_s)\epsilon_0 \tag{2}$$

$$= -m\epsilon_0 \tag{3}$$

Here, we measure the energy of the system by the unit of energy ϵ_0. The quantities w_l and w_s stand for the relative weight of long- and short-range interactions, respectively. We note here that the quantities m_l, m_s, and m can be used as order parameters of the system.

Amino Acid Substituted Models

It has been recognized that various energy terms operative in the native state of real proteins are generally consistent with each other and therefore consistent with the native conformations. One of the authors has proposed to recognize this fact as the principle of consistency among various energy

terms.[4] The intramolecular interactions assumed above in the Original model satisfy this principle.

Concerning the influence of amino acid substitution on intramolecular interactions, we assume that the substitution affects only the short-range interactions, making them locally inconsistent with the long-range interactions. This would be a natural assumption for mutated proteins. We adopt two substituted models. Substitution 1, which has unfavorable short-range interactions at the 39- and 40th units (shown in Fig. 3 as hatched squares). Substitution 2 has unfavorable ones at the 22- and 23th units (also shown in Fig. 3 as hatched squares). The difference between these two models is the location of amino acid substitution. The location in Substitution 1 is in the middle of the β-sheet region. The one in Substitution 2 is at the junction between one of the α-helices and the β-sheet. The two substituted model proteins are still expected to take the native conformation of Fig. 1, since the perturbational effect is small. If this is the case, they have native states with small amounts of local strain. We are interested to see the effects of this strain on the behavior of the protein during the folding and unfolding transition, especially on its kinetic aspects.

Method of Simulation

The simulation of the conformational transition of the lattice protein is done by the Monte Carlo method of Metropolis et al.[12] In this method, first the *a priori* transition probability p_{ij}^0 from conformational state i to j is assumed. Then, the real transition probability p_{ij} is determined by three factors; conformational enthalpy H_i of state i, H_j of state j, and the temperature of the system T, as follows.

$$p_{ij} = \begin{cases} p_{ij}^0, & \text{for } H_i \geqq H_j \\ p_{ij}^0 \exp\left[(H_i - H_j)/kT\right], & \text{for } H_i < H_j \end{cases} \tag{4}$$

where k is the Boltzmann constant. The *a priori* transition probability is taken to satisfy the condition of detailed balance,

$$p_{ij}^0 = p_{ji}^0 \tag{5}$$

Metropolis et al. demonstrated that, in a long run of the simulation, an individual state appear with the probability given by the Boltzmann distribution law.[12] We note here that the temperature of the system T comes into this method in the form of Eq. (4).

It was shown in a previous paper[5] that the number of attempted steps in this method can be regarded to be approximately proportional to physical time, if the transition probability is chosen properly. Hence, our simulation is regarded as chronological events where the lattice protein fluctuates stochastically between native (N) and denatured (D) states. Details of the method of simulation have been described in earlier papers.[5,8,9]

Analysis of Record of Simulation

By carrying out a long-enough simulation, we get a series of events of conformational transitions, which we monitor by the order parameters m_1, m_8, m, etc., introduced above. We analyze the results from both equilibrium (thermodynamic) and kinetic aspects. Hereafter, we use the dimensionless temperature T^*, defined by kT/ϵ_0, where ϵ_0 is the unit of energy introduced in Eqs. (2) and (3).

Equilibrium Analysis

Analysis based on equilibrium constant. If the transition is highly cooperative and proceeds as a two-state (all-or-none) process between N and D states, the equilibrium constant of the system $K(T^*)$ at each temperature is given by

$$K(T^*) = \mathrm{D}(T^*)/\mathrm{N}(T^*) \tag{6}$$

$$= \exp(-\Delta G/T^*) \tag{7}$$

where $\mathrm{D}(T^*)$ and $\mathrm{N}(T^*)$ are the concentrations of denatured and native states, respectively. ΔG is the difference of free energy between the two states measured by the unit of energy ϵ_0, defined by $\Delta G = G_\mathrm{D} - G_\mathrm{N}$. Rewriting Eq. (7) as

$$\Delta G = -T^* \ln K(T^*) \tag{8}$$

we can derive various kinds of equilibrium quantities by the usual thermodynamic relations, once $K(T^*)$ is determined. For example, the difference of enthalpy ΔH (by the unit of energy ϵ_0) and entropy ΔS (by the unit of the Boltzmann constant k) of the system are given by

$$\Delta H = H_\mathrm{D} - H_\mathrm{N} \tag{9}$$

$$= -\partial [\ln K(T^*)]/\partial(1/T^*) \tag{10}$$

$$\Delta S = S_\mathrm{D} - S_\mathrm{N} \tag{11}$$

$$= -\partial \Delta G/\partial T^* \tag{12}$$

Analysis based on the population profile[8]. According to Eq. (3), the energy (enthalpy) of the system H_m characterized by energy count m is given by

$$H_m = -m\epsilon_0 \tag{13}$$

Similarly, the entropy of the system S_m characterized by energy count m is defined theoretically by

$$S_m = k \ln \Omega_m \tag{14}$$

where k is the Boltzmann constant and Ω_m is the number of all possible

different conformations whose conformational energy is $-m\epsilon_0$. Note that S_m is a temperature-independent quantity. From Eqs. (13) and (14), the probability of the system p_m to be in a state characterized by energy count m is given by

$$\begin{aligned} p_m &= c\Omega_m \exp(-H_m/kT) \\ &= c \exp(S_m/k + m/T^*) \end{aligned} \tag{15}$$

where c is a proportionality constant. Rewriting Eq. (15), we get the following relation:

$$S_m/k = \ln p_m - m/T^* + \text{const} \tag{16}$$

We use this equation to determine S_m/k from the "experimental" value of probability p_m (we call it "population profile"), which can be obtained from simulation data by counting the frequency to be in a state characterized by energy count m. Practically, S_m/k is best determined from one long record of simulation carried out at or near the transition temperature.

Once temperature-independent S_m/k is known, the population profile at any temperature can be calculated by using Eq. (15). The equilibrium transition curve is then obtained from the population profile by

$$\theta = \sum_m m p_m / m_{\max} \tag{17}$$

where $m_{\max}$ is the maximum value of energy count m. The specific heat, related to the energy fluctuation of the system, is given by

$$c/k = \langle (m - \langle m \rangle)^2 \rangle / T^{*2} \tag{18}$$

Kinetic Analysis

We define the normalized time correlation function $\phi_p(t)$ of the order parameter $p(t)$ from a long simulation record as follows[13]:

$$\phi_p(t) = (\langle p(0)p(t) \rangle - \langle p \rangle^2)/(\langle p^2 \rangle - \langle p \rangle^2) \tag{19}$$

As order parameter p, we select, first, the counts of long-range order m_1. Second, we introduce the radius of gyration s defined by

$$s = \left[\sum_{i=1}^{N} (\mathbf{r}_i - \mathbf{r}_G)^2 / N \right]^{1/2} \tag{20}$$

where

$$\mathbf{r}_G = \sum_{i=1}^{N} \mathbf{r}_i / N \tag{21}$$

Here N is the number of units in the lattice protein, $\mathbf{r}_i$ is the positional vector of the ith unit, and $\mathbf{r}_G$ is the center of gravity of the lattice protein. This quantity s indicates the compactness of the conformation.[13]

As shown in the next section, the time correlation function $\phi_p(t)$ can be approximated by a sum of two exponential terms.[13] We call the two, fast and slow modes, each characterized by the corresponding relaxation time, τ_f and τ_s, respectively.

$$\phi_p(t) = \alpha_f \exp(-t/\tau_f) + \alpha_s \exp(-t/\tau_s) \tag{22}$$

where α_f and α_s are relaxation amplitudes of respective modes. In the next section, it is shown that the slow mode with τ_s corresponds to the process of overall folding and unfolding.[13]

If the transition is of the two-state type, then we can treat the process by the following scheme:

$$\mathrm{N} \underset{k_f}{\overset{k_{uf}}{\rightleftharpoons}} \mathrm{D} \tag{23}$$

Here we denote the rate constant from the N to D state as k_{uf}, and the reverse one as k_f. Then, equilibrium constant K in Eq. (6) and slow relaxation time τ_s in Eq. (22) are given in terms of these rate constants by

$$K = k_{uf}/k_f \tag{24}$$

$$1/\tau_s = k_{uf} + k_f \tag{25}$$

From these equations, we have

$$k_f = 1/[(1 + K)\tau_s] \tag{26}$$

$$k_{uf} = K/[(1 + K)\tau_s] \tag{27}$$

By these relations, we are able to derive the rate constants at each temperature from values of K and τ_s which are experimentally determined by the simulation.

According to the absolute rate theory,[14] the temperature dependence of the rate constants is given by

$$k_{f(uf)} = (kT/h) \exp[-G^{\ddagger}_{f(uf)}/kT] \tag{28}$$

where (uf) in the subscripts indicates the corresponding equation, h is the Planck constant, and $G^{\ddagger}_{f(uf)}$ is the free energy of the activated state in the transition from the D to N state (the N to D state). Here, we introduce a time unit, defined by h/ϵ_0, and we measure the time in this unit. Hence, we redefine the rate constants in this unit of time. At the same time, we measure energies and free energies by the unit of energy ϵ_0. But we retain the same notation for redefined quantities for simplicity.

$$k_{f(uf)} = T^* \exp(-G^{\ddagger}_{f(uf)}/T^*) \tag{29}$$

$$= T^* \exp[S^{\ddagger}_{f(uf)}] \exp[-H^{\ddagger}_{f(uf)}/T^*] \tag{30}$$

where $H^{\ddagger}$ and $S^{\ddagger}$ are the enthalpy (by the unit of energy ϵ_0) and entropy (by the unit of the Boltzmann constant k) of activation, respectively. These quantities $H^{\ddagger}$ and $S^{\ddagger}$ are given by relations similar to Eqs. (10) and (12):

$$H^{\ddagger}_{f(uf)} = -\partial\{\ln[k_{f(uf)}/T^*]\}/\partial(1/T^*) \tag{31}$$

$$S^{\ddagger}_{f(uf)} = -\partial G^{\ddagger}_{f(uf)}/\partial T^* \tag{32}$$

Equation (31) means that $H^{\ddagger}$ is given by the slope of tangent to the plot of $\ln(k/T^*)$ against $1/T^*$. This plot is similar to the Arrhenius plot, in which $\ln k$ is plotted against $1/T^*$. From the Arrhenius plot, activation energy E^A is given by the slope of its tangent, i.e.,

$$E^{A}_{f(uf)} = -\partial(\ln k_{f(uf)})/\partial(1/T^*) \tag{33}$$

The activation energy $E^{A}_{f(uf)}$ is different from $H^{\ddagger}_{f(uf)}$ only by a small amount, that is, $H^{\ddagger}_{f(uf)} = E^{A}_{f(uf)} - T^*$. Therefore, we will use both of these plots conveniently and more or less interchangeably in this paper.

As seen from Eqs. (7), (9), (11) and (24), (29), (30), these quantities obtained by equilibrium and kinetic analysis are related to each other by

$$\Delta G = G_D - G_N = G^{\ddagger}_{uf} - G^{\ddagger}_{f} \tag{34}$$

$$\Delta H = H_D - H_N = H^{\ddagger}_{uf} - H^{\ddagger}_{f} \tag{35}$$

$$\Delta S = S_D - S_N = S^{\ddagger}_{uf} - S^{\ddagger}_{f} \tag{36}$$

These relations are represented in Fig. 4 schematically in the case of the free energy of activation. These relations are consequences of the assumption that the transition is of the two-state type both thermodynamically and kinetically.

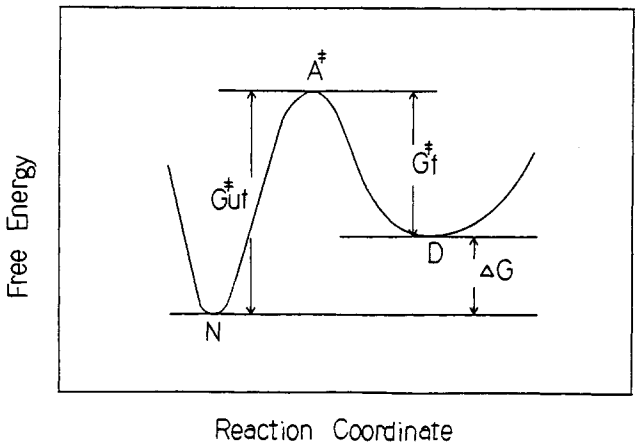

Fig. 4. Schematic representation of the relation between equilibrium and kinetic quantities for free energy of activation.

RESULTS AND DISCUSSION

Throughout this paper, we treated the case where the relative weights of long- and short-range interactions are 3 to 1, that is, $(w_l, w_s) = (0.75, 0.25)$. These relative weights were found in a previous paper[8,13] to correspond best to real proteins. If the short-range interactions are absent, the transition becomes very slow. If they are too much weighted, the transition is no more of the two-state type. In the case of the relative weights of 3 to 1, the energies of the system in the native state of Original, Substitution 1, and Substitution 2 are $-38.75\epsilon_0$, $-38.25\epsilon_0$, and $-38.25\epsilon_0$, respectively. Hence, the native state of both substituted models became unstable in energy as compared to the Original by 1.3%. This was caused by the slightly unfavorable short-range interactions.

Simulation Profile

Overall simulation profiles at temperatures in the transition zone are summarized in Tables IA, IB, and IC, for the models of Original, Substitution 1, and Substitution 2, respectively. The simulation at each temperature for each model was carried out up to 1.2×10^7 attempted steps (trial number) in most cases. Especially near the transition temperature, the simulation was repeated a few times by using a different series of random numbers in order to obtain good statistics. This was necessary because of large-scale fluctuations.

The transition temperatures T_m^* determined for the three models are shown by the arrows in Tables IA–IC. They were determined from the simulation profiles as the temperatures where the following condition is satisfied; The mean value of long-range energy count $\langle m_l \rangle$ in Tables IA–IC is around the

TABLE IA
Simulation Profile for Original Model[a]

T^*	Trials ($\times 10^{-7}$)	N ⇌ D	$\langle m_l \rangle$	$\langle s \rangle$	K	k_f ($\times 10^5$)	k_{uf} ($\times 10^5$)
0.615	1.2	30	32.8	3.14	0.130	2.91	0.38
0.630	0.9	35	30.4	3.37	0.282	2.29	0.64
0.640	1.2	42	27.2	3.67	0.543	1.35	0.73
0.650	1.2	47	25.9	3.79	0.670	1.29	0.87
0.660	1.2	54	24.4	3.94	0.930	1.19	1.11
→ 0.660	1.2	51.5	23.7	4.00	1.003	1.17	1.17
0.665	1.2	50	22.5	4.12	1.238	0.98	1.22
0.665	1.2	55	21.6	4.21	1.444	1.01	1.46
0.670	1.2	53	21.1	4.26	1.515	0.89	1.35
0.680	1.2	46.5	18.3	4.51	2.643	0.71	1.89
0.690	1.2	48.5	18.2	4.54	2.546	0.79	2.01
0.710	0.9	22	13.4	4.99	7.575	0.48	3.64
0.730	1.2	15.5	11.8	5.13	13.929	0.37	5.14

[a] Meaning of column heads: T^*, temperature; Trials, trial number; N ⇌ D, the number of times going forth and back between N and D states; $\langle m_l \rangle$, the mean of long-range energy count; $\langle s \rangle$, the means of radius of gyration; K, equilibrium constant; k_f, rate constant for folding transition; k_{uf}, rate constant for unfolding transition. Transition temperature (T_m^*) is indicated by an arrow.

TABLE IB
Simulation Profile for Substitution 1 Model[a]

T^*	Trials ($\times 10^{-7}$)	N $\rightleftharpoons$ D	$\langle m_1 \rangle$	$\langle s \rangle$	K	k_f ($\times 10^5$)	k_{uf} ($\times 10^5$)
0.615	1.2	29	32.5	3.17	0.154	2.33	0.36
0.630	1.2	36	28.7	3.53	0.413	1.26	0.52
0.640	1.2	36.5	27.4	3.66	0.532	1.18	0.63
0.645	1.2	41.5	24.6	3.93	0.854	1.01	0.87
→ 0.650	1.2	45	23.8	3.99	0.980	0.94	0.92
0.650	1.2	44	24.3	3.95	0.880	0.96	0.85
0.655	1.2	44.5	21.2	4.25	1.563	0.80	1.26
0.655	1.2	44	21.4	4.23	1.476	0.75	1.10
0.660	1.2	45	20.9	4.27	1.576	0.76	1.20
0.670	0.9	29.5	18.4	4.52	2.353	0.65	1.54
0.690	1.2	22.5	13.7	4.95	7.648	0.34	2.61
0.710	1.05	20	12.6	5.06	8.714	0.38	3.35

[a] Meaning of column heads same as in Table IA. Transition temperature (T_m^*) is indicated by an arrow.

TABLE IC
Simulation Profile for Substitution 2 Model[a]

T^*	Trials ($\times 10^{-7}$)	N $\rightleftharpoons$ D	$\langle m_1 \rangle$	$\langle s \rangle$	K	k_f ($\times 10^5$)	k_{uf} ($\times 10^5$)
0.600	1.2	25	33.7	3.06	0.090	4.34	0.39
0.615	1.2	37	30.7	3.34	0.275	1.75	0.48
0.620	1.2	39	30.7	3.34	0.264	1.91	0.50
0.630	1.2	49	27.3	3.67	0.570	1.41	0.80
0.630	1.2	48	27.9	3.61	0.506	1.49	0.76
0.635	1.2	45	27.6	3.64	0.528	1.53	0.81
0.640	1.2	52	26.5	3.75	0.649	1.36	0.88
→ 0.645	1.2	43.5	24.3	3.96	0.986	1.04	1.02
0.645	1.2	51.5	24.9	3.91	0.880	1.11	0.98
0.650	1.2	47.5	23.2	4.07	1.184	0.94	1.11
0.660	0.9	33.5	21.0	4.28	1.727	0.89	1.54
0.670	0.9	21.5	17.4	4.64	3.516	0.55	1.94
0.690	1.2	30.5	14.9	4.87	6.012	0.51	3.06
0.710	1.05	17.5	12.4	5.10	11.975	0.34	4.07

[a] Meaning of column heads same as in Table IA. Transition temperature (T_m^*) is indicated by an arrow.

value of 24, which is the expected value at the transition temperature, i.e., the mean of the maximum value 36 ($m_1^{\max}$ in the N state) and the minimum value of about 12 (the highest temperature side in Tables IA–IC).

At the transition temperature, the value of $\langle s \rangle$ is about 4.0. This is the value expected at the transition temperature, i.e., the mean of $2\sqrt{2} \fallingdotseq 2.8$ (the N state in Fig. 1) and about 5.1 (the highest temperature side in Tables IA–IC).

The values of T_m^* thus determined for the models of Original, Substitution 1, and Substitution 2 are 0.660, 0.650, and 0.645, respectively. As the accuracy of the determination of T_m^* may be within ± 0.005 (the minimal step width of temperature in the simulation), the difference between the values of T_m^* for

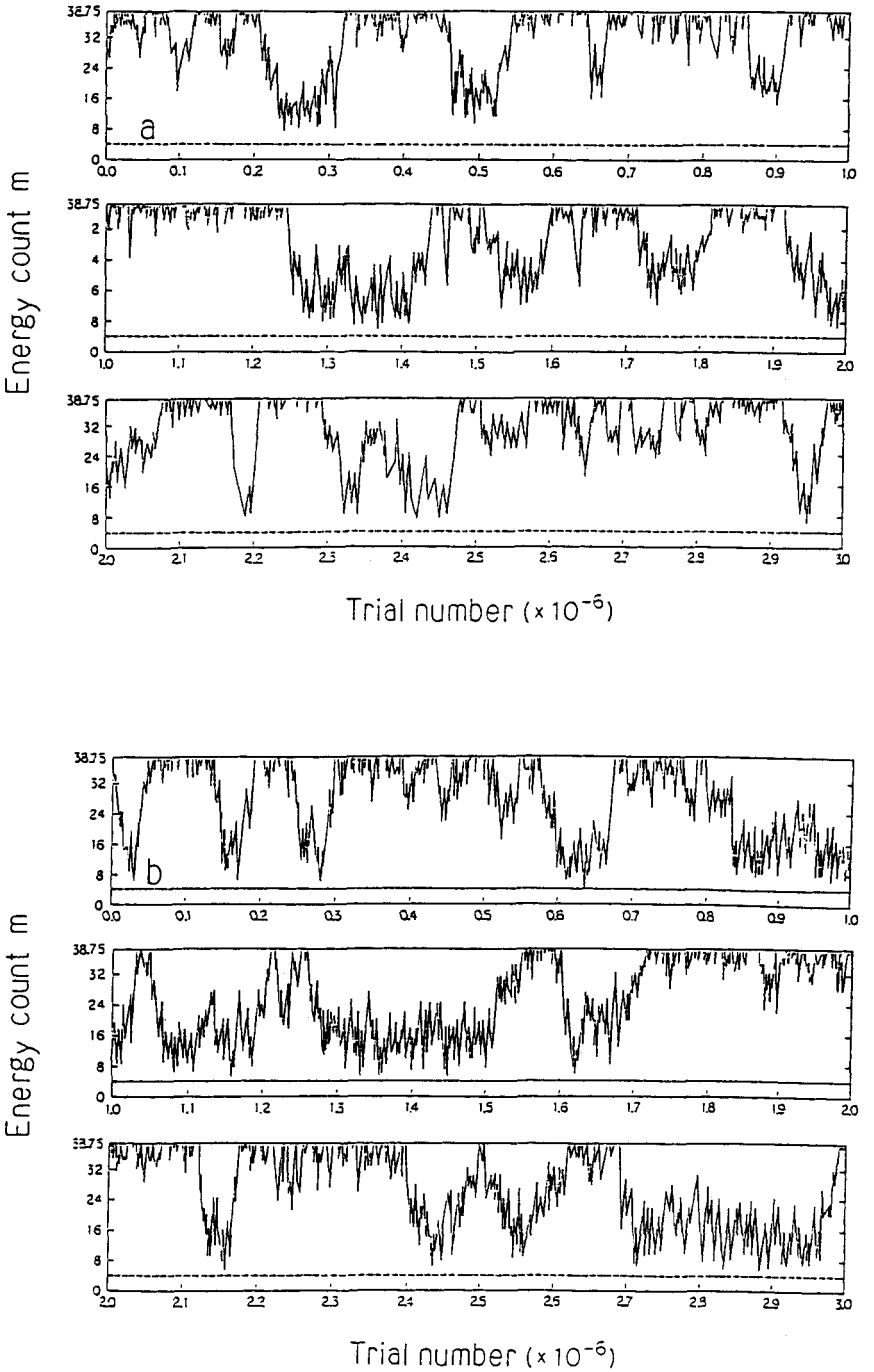

Fig. 5(a). Time course of long simulation run monitored by energy count m for Original model at $T^* = 0.630$ ($< T_m^*$). Abscissa and ordinate indicate the number of attempted steps (trial number) and the energy count m, respectively. (b) Same as in (a), but for $T^* = 0.660$ ($= T_m^*$). (c) Same as in (a), but for $T^* = 0.690$ ($> T_m^*$).

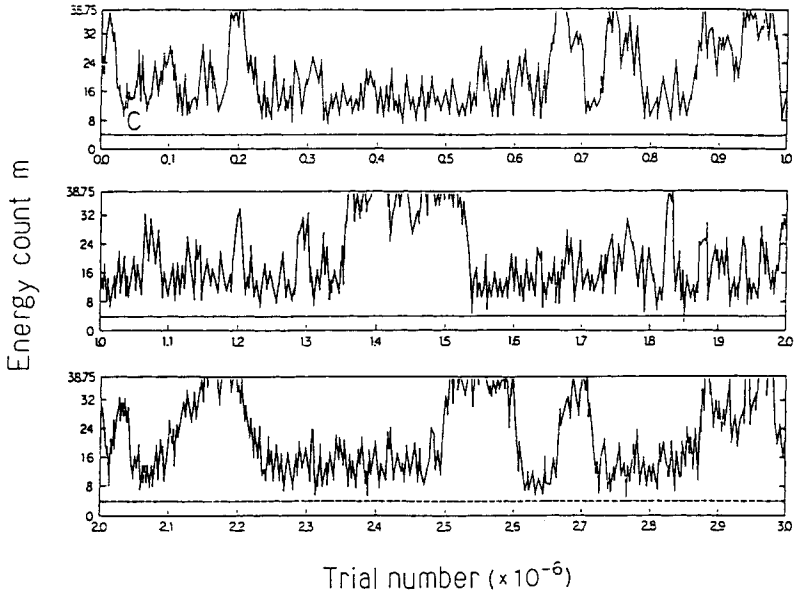

Fig. 5. *(Continued from the previous page.)*

Substitution 1 and Substitution 2 is within the computational inaccuracy. However, it is certain that T_m^* for the two substituted models became lower than the one for Original model.

Figure 5 shows the representative time courses of a long simulation run for the Original model as monitored by energy count m (we call them "energy profiles") for three cases (a) $T^* = 0.630$ ($< T_m^*$), (b) 0.660 ($= T_m^*$), and (c) 0.690 ($> T_m^*$). The time courses for the other substituted models are similar. The broken lines at $m = 0.25 \times 47 \times (1/3)$ indicate the mean of the short-range energy count at a high temperature limit.

It should be noted, first, that the lattice protein stays most of the time at either the N or D state, and occasionally rapid transitions between the two states take place. This means that the transition is highly cooperative and of two-state (all-or-none) type. Second, the manner in which a transition occurs, especially in the folding transition, is quite different from case to case, indicating the stochastic nature of the transition. Sometimes the intermediate states with fairly high values of order parameter do not return to the native state directly, but are destroyed. This suggests that there is a nucleation process in the folding transition and it may be a rate-limiting process of the transition. By nucleation process we mean a process whereby a nucleus of a critical size is formed, which, once formed, tends to grow further to lead the protein molecule to the native state. Such a critical state may be the activated state of the transition, as shown in Fig. 4. Third, the transition becomes more frequent as temperature approaches T_m^*, although above T_m^* it becomes infrequent again. Fourth, considerable residual structure remains in the D state, as shown by the values of $\langle m_1 \rangle$ in Table I. This is one of the characteristics of thermal transition in real proteins.[1,2]

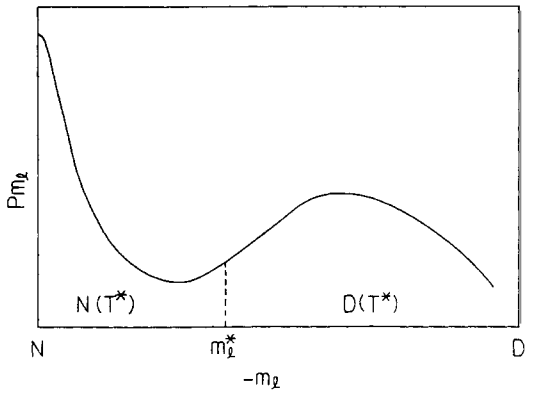

Fig. 6. Schematic representation of the population profile for long-range energy count m_1. m_1^* is the mean value of m_1, that is, $\langle m_1 \rangle$, at $T^* = T_m^*$.

Equilibrium Properties

Analysis by Equilibrium Constant

The value of equilibrium constant K at each temperature for each model is tabulated in Tables IA–IC. It is determined as follows: We first calculate the population profile of m_1 at the transition temperature. From the profile we calculate $\langle m_1 \rangle$ (the mean value of m_1 at $T^* = T_m^*$). This value m_1^*, shown schematically in Fig. 6, is the point that divides the population profile into two equal parts, because the sum of the population of each side is almost equal. Then, at each temperature we calculate the area of the right and left sides of m_1^* (that is, the sum of the probability in each side), as $D(T^*)$ (the concentration of the D state) and $N(T^*)$ (the one of the N state), respectively. Equilibrium constant $K(T^*)$ is obtained as the ratio of the concentrations at each temperature by Eq. (6)

Figure 7 shows the plot of the logarithm of equilibrium constant ($\ln K$) vs $1/T^*$, that is, the van't Hoff plot for the Original model. The plot can be fitted to a good approximation by a straight line, as shown by the solid line in Fig. 7. The slope gives the difference of enthalpy ΔH [Eq. (10)]. The value of ΔH is calculated by the least-square fitting and tabulated in Table II. The ones for the substituted models are also given. In Fig. 8, the fitted lines of the van't Hoff plots for the three models are shown. It indicates that the amino acid substitutions cause the substituted models to be slightly unstable as seen from the shift of the lines from the Original model.[15,16] This explains the decrease of T_m^* for the substituted models that were noted earlier.

The linearity of the van't Hoff plot means that the plot of ΔG vs T^* is also linear. Thus, the slope gives the value of ΔS [Eq. (12)]. The values thus determined are also tabulated in Table II. Note here that the linearity of the plot of ΔG vs T^* differs clearly from the real experimental data. The latter shows a maximum below the transition temperature.[17-20] The following two reasons may explain these discrepancies: First, hydrophobic interactions are neglected in the theoretical model treated in this study. Second, the above

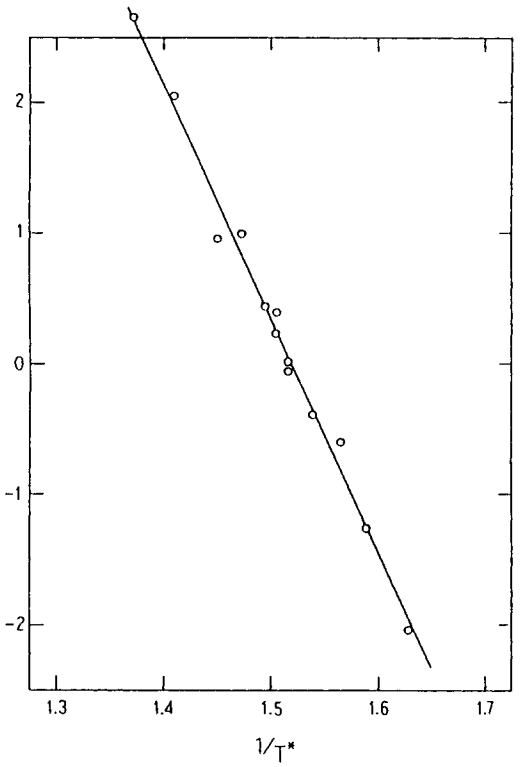

Fig. 7. van't Hoff plot (ln K vs $1/T^*$) for Original model. The slope gives the value of the difference of enthalpy ΔH.

TABLE II
Thermodynamic Parameters for the Three Models

	ΔH	ΔS	$T_m^* = \Delta H/\Delta S$
Original	17.9	27.1	0.659
Substitution 1	19.4	29.6	0.655
Substitution 2	18.2	28.3	0.645

result for the theoretical model has been obtained from simulations in the transition zone, although the maximum in real experimental data occurs at temperatures fairly well below the transition temperature.[17-20]

The transition temperature T_m^* is also defined in terms of ΔH and ΔS determined above by

$$T_m^* = \Delta H/\Delta S \tag{37}$$

The values of T_m^* thus calculated for the three models are tabulated in Table II. These values reproduce those that were determined in the preceding subsection, except for a slight deviation in Substitution 1. This good reproduction ensures the validity of the equilibrium analysis based on equilibrium constants.

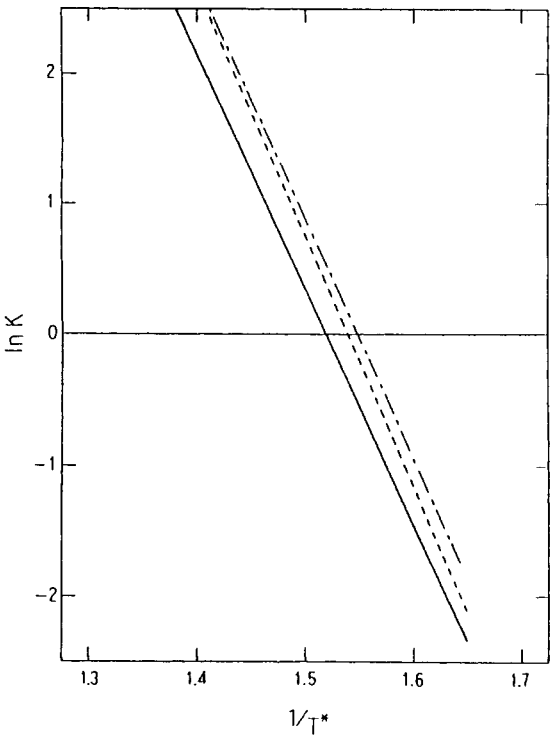

Fig. 8. The fitted lines of van't Hoff plot for the three models. (——) Original, (- - - - -) Substitution 1, and (—·—·—) Substitution 2.

Analysis by Population Profile

Here we present only the results for the Original model, because the results for other substituted models are qualitatively similar. Figure 9 shows the population profile p_m obtained from the simulation at the transition temperature, $T^* = T_m^*$. It shows that the distribution has nearly double maxima, indicating the two-state character of the transition. But the denatured states

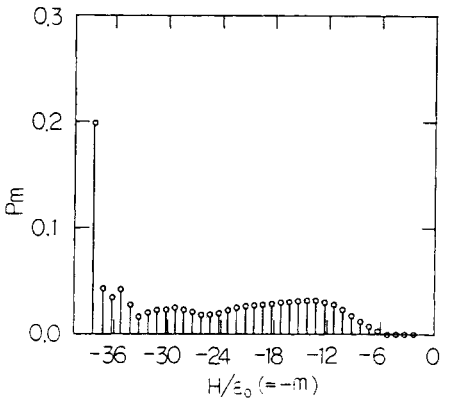

Fig. 9. Population profile P_m of energy count m for the Original model at $T^* = T_m^*$.

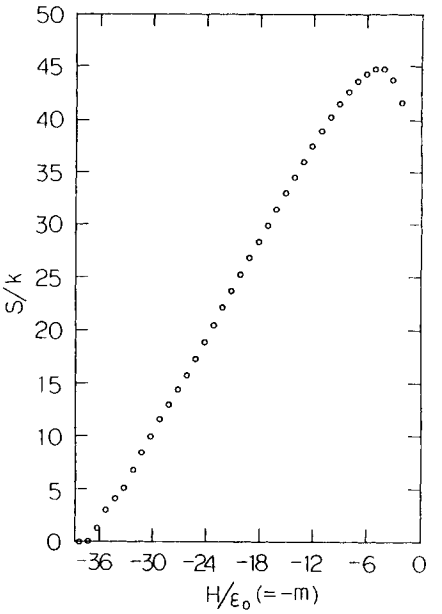

Fig. 10. S-H curve for Original model. The concavity of the curve indicates the all-or-none type transition.

are widely spread. From the population profile, the S-H curve introduced previously[21] is obtained by Eq. (16) and shown in Fig. 10. The concavity of the curve is a good index of two-state (all-or-none) type transition.[21] In Fig. 11, the equilibrium transition curve θ is shown, which is obtained by Eqs. (15) and (17). The residual structures can be clearly observed in D states as the

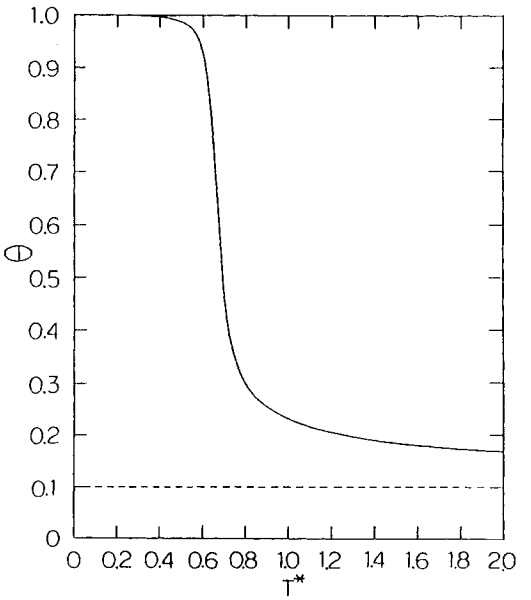

Fig. 11. Equilibrium transition curve θ vs T^* for the Original model.

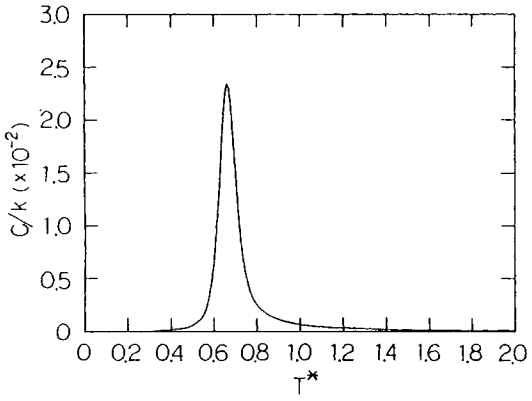

Fig. 12. Specific heat of the system c/k vs T^* for the Original model.

fairly large value of θ at the high temperature side. Figure 12 shows the curve of specific heat c/k vs temperature, which is obtained by Eq. (18). It shows a sharp peak near the transition temperature T_m^*.[17,18]

Relaxation Time

In kinetic analysis, we first calculate the normalized time correlation function $\phi_p(t)$, defined by Eq. (19), from the long-time simulation records for the order parameters m_1 and s. The function thus calculated is shown in Fig. 13 for the long-range energy count m_1 in the Original model at three different temperatures. It shows that the slope of the plot becomes gentler near the transition temperature T_m^*, indicating the slowing down of the relaxation

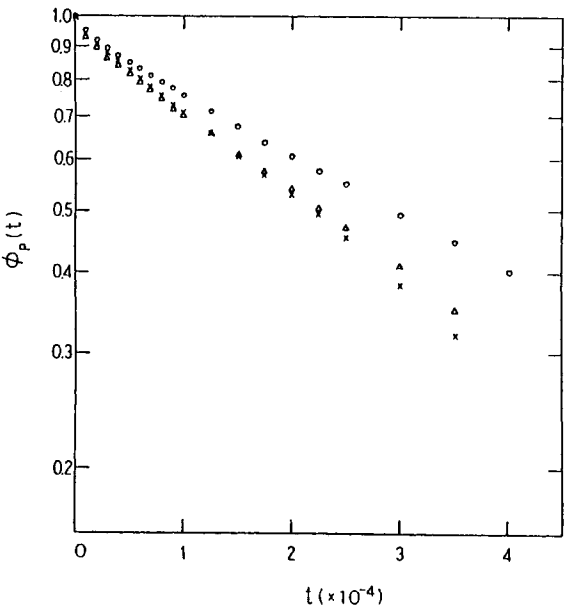

Fig. 13. Calculated time correlation function $\phi_p(t)$ for the order parameter m_1 for the Original model. (×) $T^* = 0.630$ ($< T_m^*$), (○) $T^* = 0.660$ ($= T_m^*$), (△) $T^* = 0.690$ ($> T_m^*$).

process. The plot can be fitted to a good approximation by a sum of two exponential terms, that is, by the expression of Eq. (22). As mentioned before in the previous section, we call the two phases fast and slow modes, which are characterized by the corresponding relaxation times. The fitting was carried out by using a program system SALS for nonlinear least square fitting.[22] From the analysis, relaxation times and amplitudes of each mode were obtained.

The temperature dependence of the slow relaxation time τ_s and the fast one τ_f in the transition zone is shown in Fig. 14(a–c) for the three models, respectively. The values of τ_s in Fig. 14 for the slow mode for order parameters m_1 and s agree with each other quite well. This agreement of τ_s, despite the different nature of the order parameters, indicates that the slow mode corresponds to the process of overall folding and unfolding transition. Note that, for all three models, τ_s has remarkable temperature dependence, showing a peak near the transition temperature T_m^*. It is a quantitative expression of the slowing down of the relaxation process near T_m^*. However, the emphasis should be placed on the fact that, for each model, the peak occurs at a slightly lower temperature than T_m^*. The reason for this phenomenon is given in the appendix, although it is unclear in case of Substitution 2.

On the other hand, the values of τ_f in Fig. 14 for the fast mode for order parameters m_1 and s differ from each other by one order of magnitude. This means that, in the fast mode, the relaxation process can be observed with different time scales, depending on the nature of the order parameters. Since the slow mode is regarded as reflecting the overall folding and unfolding transition, the fast mode should reflect the conformational fluctuations within each of the N and D states. It should be noted here that, in contrast to τ_s, τ_f for both m_1 and s does not show remarkable temperature dependence, but is almost insensitive to temperature.

Temperature dependence of the relaxation amplitudes for the order parameters m_1 and s is shown in Fig. 15 and Fig. 16, respectively, for the Original model. It is shown in Fig. 15 that the slow mode becomes dominant near T_m^* for the order parameter m_1.

A comparison between Fig. 14(a) and Fig. 14(b) shows that the slow relaxation time for Substitution 1 is longer than the one for the Original by about 20%. It indicates that the slightly unfavorable substitutions of amino acids cause the transition to slow.[23] However, the comparison between Fig. 14(a) and Fig. 14(c) does not show remarkable difference, although the relaxation time for Substitution 2 is a bit longer than the one for the Original. The two substituted models differ only in the location of amino acid substitution. Substitution 1 has two substituted units in the middle of the β-sheet region, while Substitution 2 has ones at the junction between one of the α-helices and the β-sheet. The difference of the above kinetic behavior between the substituted models should be attributed to the location of amino acid substitution. The cause of the difference of these kinetics is studied further below.

Rate Constant

Rate constants for the folding and unfolding transition for the three models are calculated at each temperature from the values of K and τ_s (average value

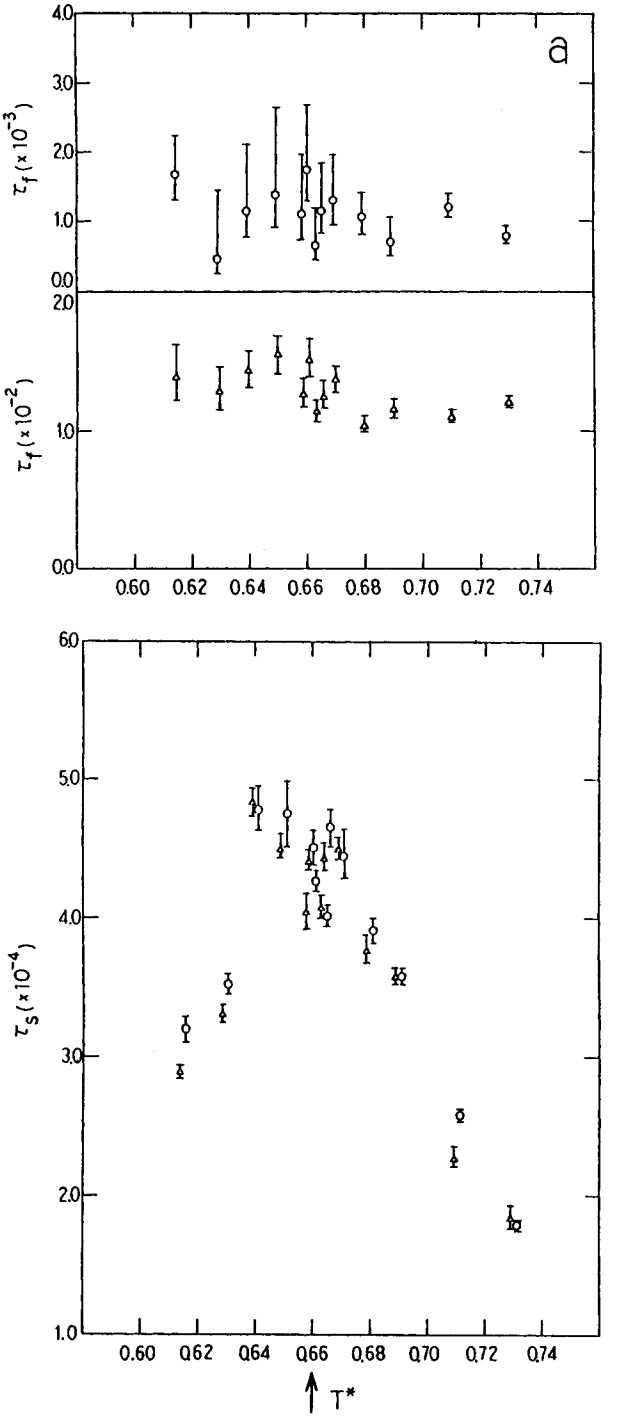

Fig. 14(a). Temperature dependence of the slow relaxation time τ_s (lower) and the fast relaxation time τ_f (upper) for order parameters m_1 (○) and s (△) for the Original model. Transition temperature T_m^* is indicated by an arrow. (b) Same as in (a), but for the Substitution 1 model. (c) Same as in (a), but for the Substitution 2 model.

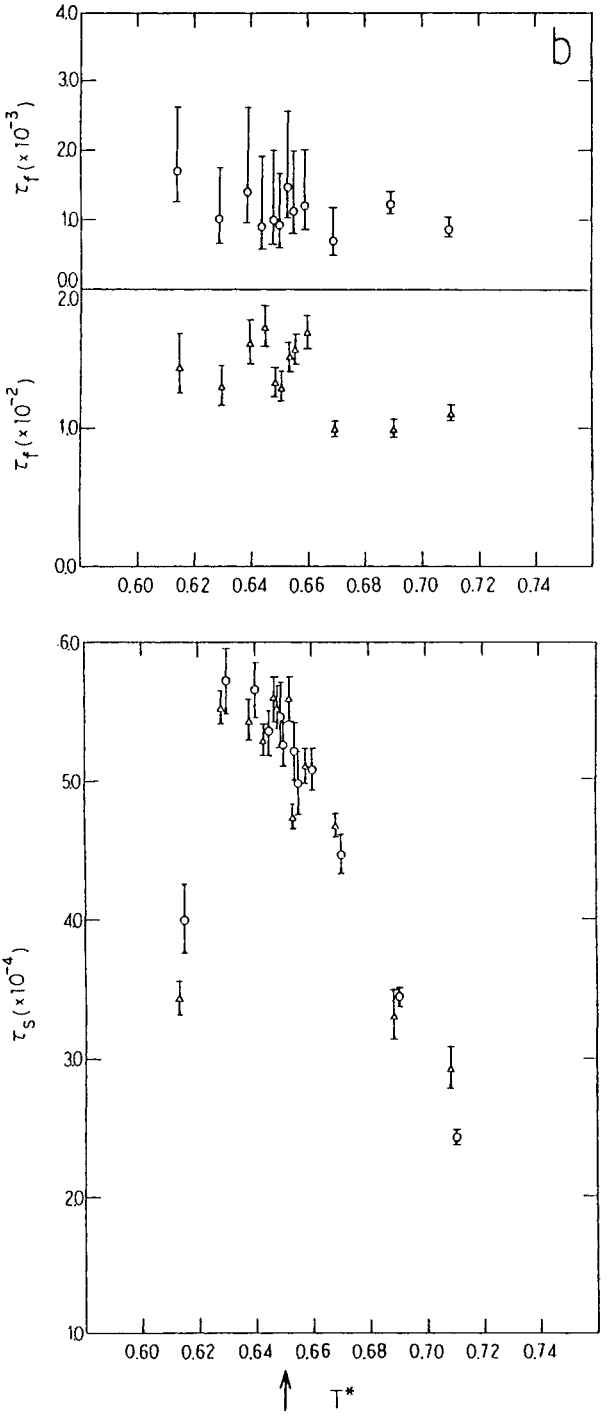

Fig. 14. (*Continued from the previous page.*)

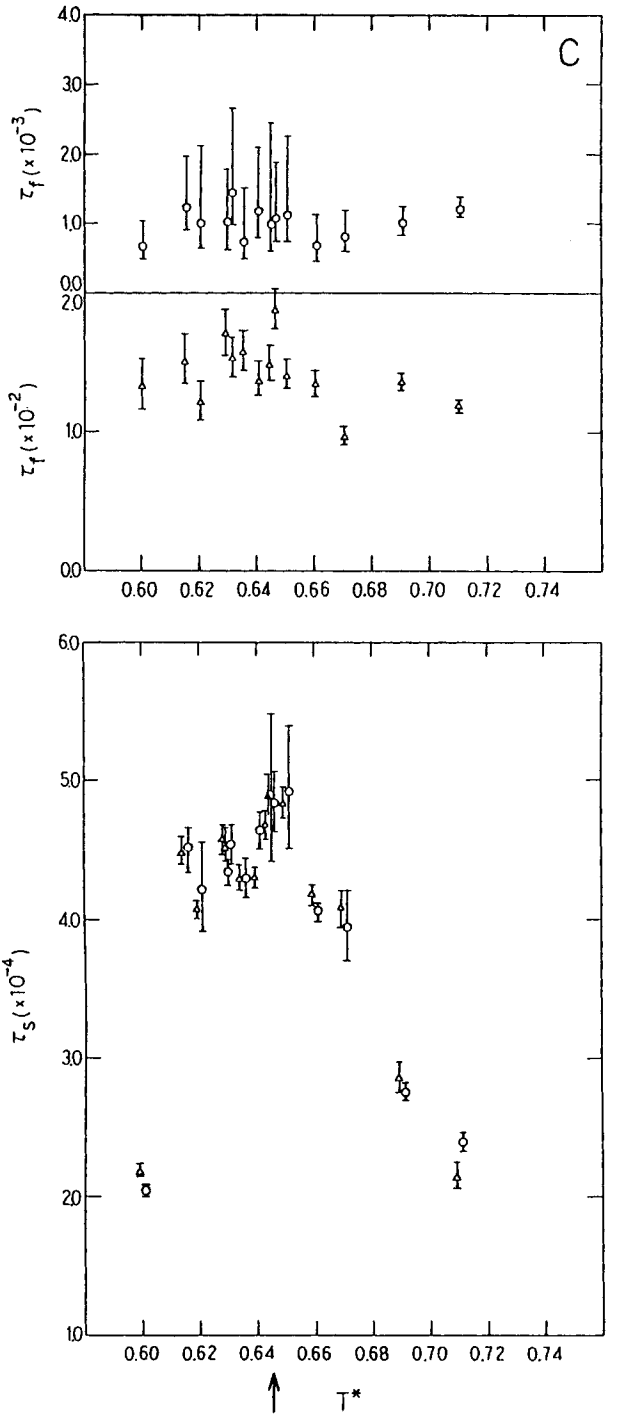

Fig. 14. *(Continued from the previous page.)*

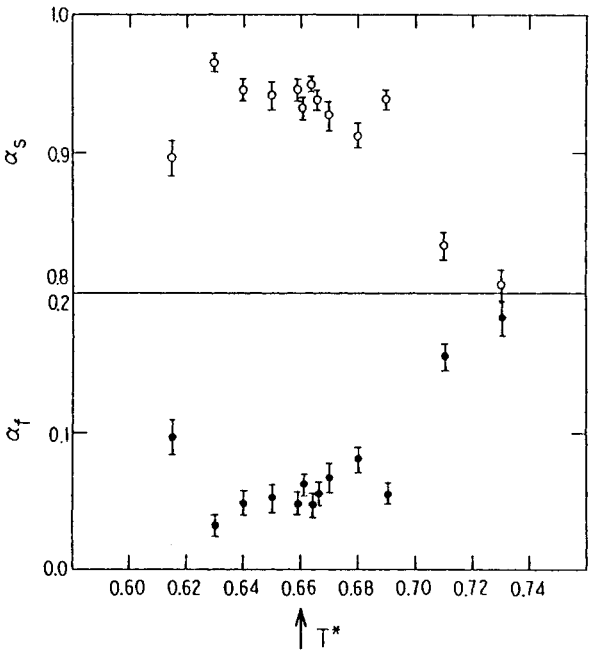

Fig. 15. Temperature dependence of the relaxation amplitude for order parameter m_1 for the Original model. Transition temperature is indicated by an arrow. Top: slow mode; bottom: fast mode.

for order parameters m_1 and s) obtained in preceding subsections by using Eqs. (26) and (27). They are tabulated in Tables IA–IC.

Figure 17 shows the plot of the logarithm of rate constants [$\ln k_{f(uf)}$] vs $1/T^*$, that is, the Arrhenius plot for the Original model. The plot can be fitted by a straight line, as shown by the solid line for the folding process and a broken line for the unfolding process. The slope gives the activation energy $E^A_{f(uf)}$ by Eq. (33). The values of $E^A_{f(uf)}$ obtained by the least-square fitting are tabulated in Table III together with the ones for the substituted models. Coupled with the Arrhenius plot, plots of $\ln(k/T^*)$ vs $1/T^*$ have also been made (not shown here). In addition, these plots can be fitted by straight lines. [Strictly speaking, a straight Arrhenius plot means a slightly curved plot of $\ln(k/T^*)$ vs $1/T^*$. However, this curvature is numerically very small.] The values of the enthalpy of activation $H^‡_{f(uf)}$ are determined from the slope of the latter plots, and are also tabulated in Table III for the three models. Two points should be noted. First, the Arrhenius plot obtained give almost linear dependence on temperature even for the folding process. Second, the slopes of the plots for unfolding are a bit larger than the ones for folding for all three models, as indicated by the values of $E^A_{f(uf)}$ or $H^‡_{f(uf)}$.

Similarly, the plot of $G^‡_{f(uf)}$ defined by Eq. (29) against T^* is shown in Fig. 18 for the Original model. The slope of this plot gives the entropy of activation $S^‡_{f(uf)}$ defined by Eq. (32). The values are tabulated in Table III together with the ones for the substituted models. Two points should be emphasized. First, the plots for folding and unfolding processes give almost linear dependence on temperature. Second, the plot for the folding process is

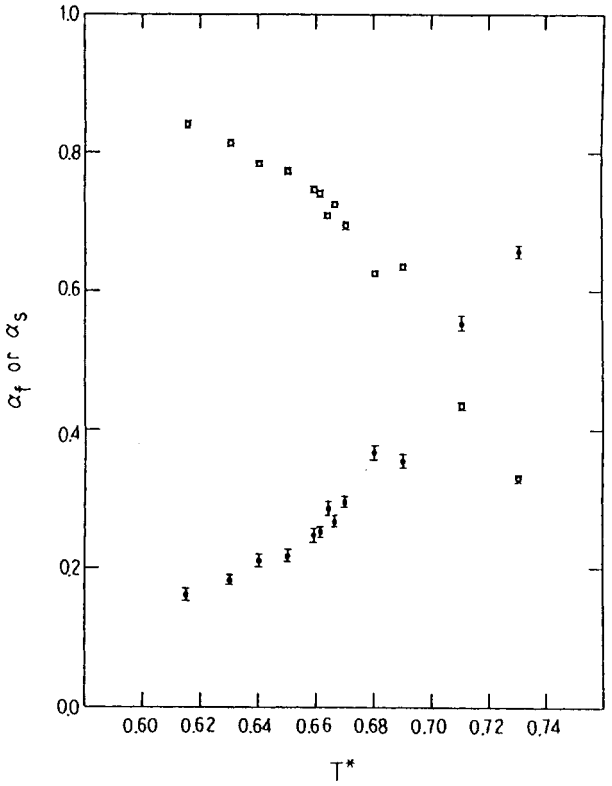

Fig. 16. Same as in Fig. 15, but for order parameter s. (○) Slow mode; (●) fast mode.

fairly dependent on temperature, while the one for the unfolding process is nearly insensitive to temperature. This is reflected in the fact that the value of $S_f^{\ddagger}$ is much larger than the one of $S_{uf}^{\ddagger}$ for all three models. This observation suggests that the entropic contribution is large for the folding process, that is, for the process from D states to the activated state, but for the unfolding process it is not, even though we are neglecting the hydrophobic interaction in our present treatment. As seen from Tables II and III, the relations of Eqs. (34)–(36) hold with a high accuracy, indicating that the transition is of all-or-none type both thermodynamically and kinetically.

The fitted lines of the plot of $G_{f(uf)}^{\ddagger}$ vs T^* for the Original and Substitution 1 models are compared in Fig. 19. It is clearly observed that $G_{uf}^{\ddagger}$ is almost the same between the two models, while $G_f^{\ddagger}$ for Substitution 1 is larger than that for the Original. This increase of the free energy of activation for the folding process makes the folding transition slower.[23] This is just the reason why Substitution 1 has a slower relaxation time than the Original. The slowing down is probably because the formation step of the β-sheet region during folding in Substitution 1, which is partially prevented by the local defects (strains) of the β-sheet region, is a rate-limiting process of the transition.

A similar comparison between the Original and Substitution 2 models is made in Fig. 20. It indicates that $G_f^{\ddagger}$ for Substitution 2 is larger than that for the Original by nearly the same amount as Substitution 1 mentioned above,

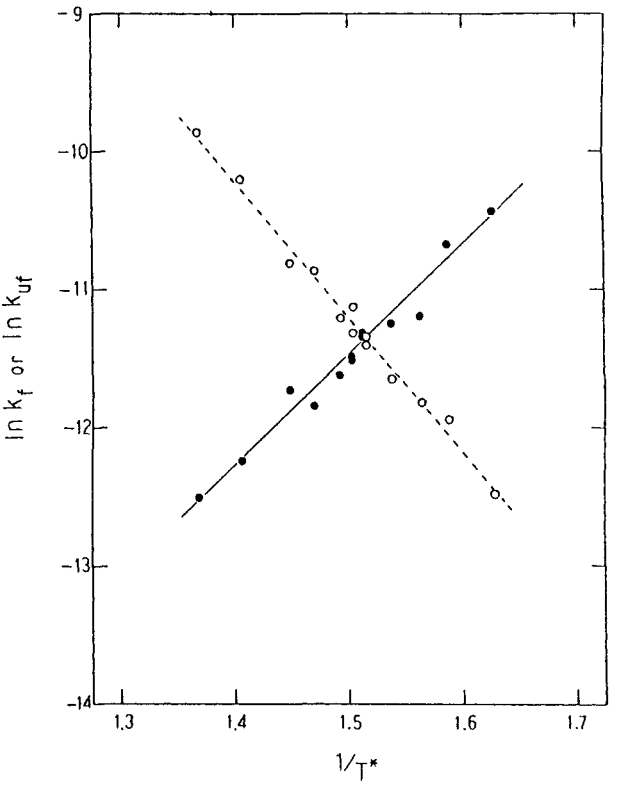

Fig. 17. Arrhenius plot (ln $k_{f(uf)}$ vs $1/T^*$) for the Original model. The slope gives the value of activation energy $E^A_{f(uf)}$. (●) Folding transition; (○) unfolding transition.

TABLE III
Kinetic Parameters of Folding and Unfolding Transitions for the Three Models

	Unfolding			Folding		
	E^A_{uf}	$H^‡_{uf}$	$S^‡_{uf}$	E^A_f	$H^‡_f$	$S^‡_f$
Original	9.9	9.3	3.2	−8.0	−8.7	−24.0
Substitution 1	10.9	10.2	4.4	−8.6	−9.3	−25.2
Substitution 2	10.1	9.5	3.6	−8.1	−8.7	−24.5

while $G^‡_{uf}$ for Substitution 2 is smaller than that for the Original. The increase of the free energy of activation for the folding process has the same effect of making the folding transition slower, as in the case of Substitution 1. However, the decrease in free energy of activation for the unfolding process makes the unfolding transition fast. Thus, these two effects tend to cancel each other when observed by the relaxation time. This is just the reason why Substitution 2 has almost the same relaxation time as the Original.

The result suggests that the β-sheet region in Substitution 1 is so rigid during unfolding that the defects there do not affect the rate of unfolding, but once it is destroyed, the defects can affect the rate of folding. However, the

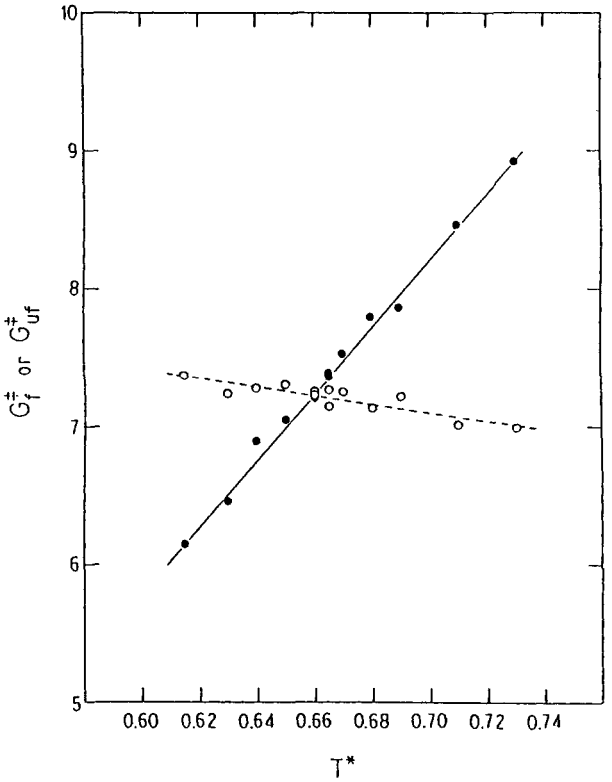

Fig. 18. The plot of free energy of activation $G_{f(uf)}^{\ddagger}$ against T^* for the Original model. The slope gives the value of entropy of activation $S_{f(uf)}^{\ddagger}$. (●) Folding transition; (○) unfolding transition.

defects in the junction part in Substitution 2 can affect the rates of both processes equally.

Activated State of the Conformational Transition

Here we make a remark on the activated state of the transition in our lattice protein.[24] As seen from Table III, the kinetic parameters $H_f^{\ddagger}$ and $H_{uf}^{\ddagger}$ are nearly similar to each other in absolute values, while $S_f^{\ddagger}$ is much larger than $S_{uf}^{\ddagger}$. This means that the activated state is located nearly halfway between the N and D states on the enthalpy axis, while it is very close to the N state on the entropy axis. Thus, we have to regard the activated state as one in which the interactions between the units are half broken, but the number of possible different conformations is quite limited.

As mentioned in the previous subsection, the local defects of the β-sheet region in Substitution 1 have an effect on the rate of the folding transition, but not on the rate of the unfolding transition. However, the defects of the junction part in Substitution 2 have effects on rates of both folding and unfolding transitions. From this difference of kinetic behavior and the activated state mentioned above, the following conclusions about the folding pathway of this lattice protein can be obtained[24]: a) The β-sheet structure is formed

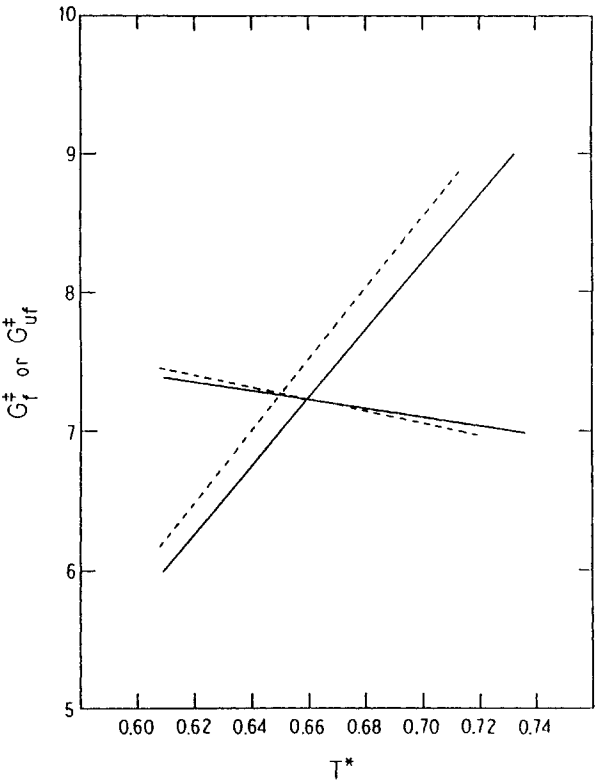

Fig. 19. The fitted lines of the plot of $G_{f(uf)}^{\ddagger}$ vs T^* for the Original and Substitution 1 models. (——) Original; (- - - - -) Substitution 1.

before it gets assembled with the α-helix structures, and (b) the process of assembly occurs nearly at the activated state of the transition.

Effect of Hydrophobic Interactions on Conformational Transition of Proteins

As mentioned in the introduction, we confine ourselves in this study to the case of no hydrophobic interactions. However, it is possible to discuss effects of the hydrophobic interactions on conformational transition of proteins to some extent by comparing the results obtained in this study with real experimental data. Comparison is made as to the following two aspects:

Temperature Dependence of Thermodynamic and Kinetic Parameters

As was shown in Fig. 7, we fitted the van't Hoff plot and hence the plot of ΔG vs T^* as linear ones. This led to the constant values of ΔH and ΔS throughout the transition zone as tabulated in Table II. However, in real small proteins, ΔH and ΔS have significant temperature dependences that increase with temperature.[17-20] This temperature dependence gives rise to a nonvanishing difference of specific heat between native and denatured states,

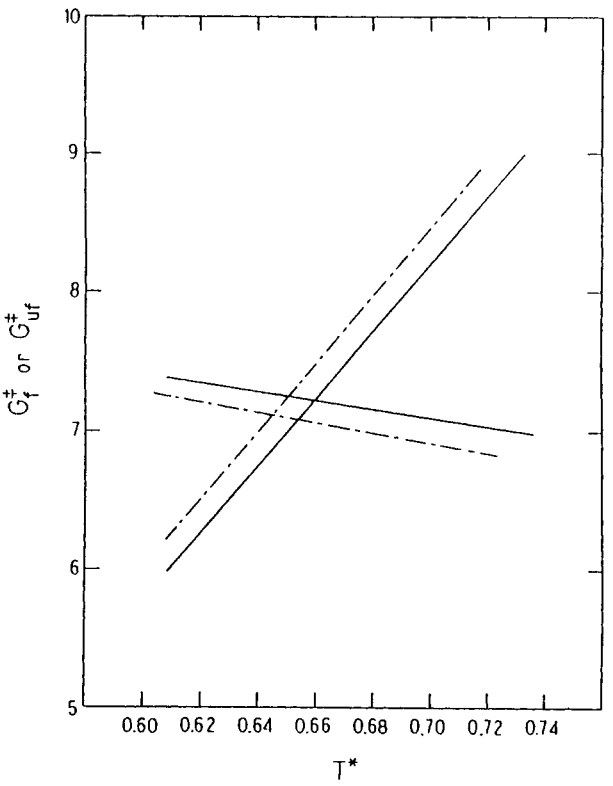

Fig. 20. Same as in Fig. 19, but for the Original and Substitution 2 models. (——) Original; (——·——) Substitution 2.

i.e., $\Delta c_p = c_{p,D} - c_{p,N} = c_{p,uf}^\ddagger - c_{p,f}^\ddagger$.[17,18] This difference between our study and experimental ones is likely due to the neglect of hydrophobic interactions in our study. In fact, in his earlier works, Brandts[25-27] interpreted a large value of Δc_p, as well as the maximum of the plot of ΔG vs T^* well below the transition temperature, as due to the exposure of hydrophobic groups during the unfolding process.

As to the kinetic parameters, we also found in this study that $G_{f(uf)}^\ddagger$ was linear with respect to temperature, as was shown in Fig. 18. In their kinetic experiments on lysozyme, Segawa and Sugihara[28,29] showed that the transition can be regarded well as a two-state process, despite the existence of the *cis–trans* isomerization reaction of Pro residues first demonstrated by Kato et al.[30,31] They also showed that the Arrhenius plot of their data for the folding process fit well with the ones by Kato et al,[31] if extrapolation is done toward lower temperatures. In the plot, the rate constant for folding k_f shows a maximum far below the transition temperature. In fact, Segawa and Sugihara[28,29] observed that $c_{p,uf}^\ddagger$ nearly equals zero and $c_{p,f}^\ddagger$ has a negative constant value. This means that kinetic parameters $G_{f(uf)}^\ddagger$ should have the following type of temperature dependence, that is, $G_{uf}^\ddagger = a + bT$ and $G_f^\ddagger = c + dT + eT*\ln T$, where $b < 0$ and $e > 0$. Thus, $G_f^\ddagger$ should have a minimum at a low temperature.

The experimentally observed nonlinear dependences of ΔG and $G_f^\ddagger$ on temperature are not independent phenomena. If the transition is of the two-state type, as is the case in lysozyme, Eq. (34), which relates an equilibrium quantity ΔG with kinetic quantities $G_f^\ddagger$ and $G_{uf}^\ddagger$, should hold. If $G_{uf}^\ddagger$ is linear to T and ΔG is nonlinear, then $G_f^\ddagger$ must be nonlinear for the same reason.

Rate Constant for Folding

As to the maximum in the plot of the folding rate constant, two types of interpretations have been proposed, i.e.,

(a) one by Lumry and Biltonen,[32] and Segawa et al.[28,29,33]; and

(b) the other by Kato et al.[31]

We discuss these interpretations from the simulation results obtained in this paper.

Lumry and Biltonen interpreted the data of reversible denaturation of chymotrypsinogen by calling for "a subtle conformational change." According to them, the process from N state to the A state (activated state) is a subtle conformational change, where the two states have about the same number of exposed hydrophobic groups, while the one from A state to D state is a gross change accompanied by the exposure of hydrophobic groups to solvent (water). Thus, in the reverse reaction, that is, during the folding process from the D state to the A state, a large negative change of the specific heat of activation caused by the hydrophobic interactions is expected, which lead to the convex curve of the Arrhenius plot for the folding process. Furthermore, they suggested that such a state with a subtle conformational change is a kind of swelling state of proteins. Segawa's interpretation of his data of lysozyme is basically the same as Lumry and Biltonen's.

On the other hand, Kato et al. interpreted their kinetic data for refolding of lysozyme as evidence of a nucleation-controlled process. From an analogy of the classical theory of the rate of nucleation in condensed systems, they concluded that a maximum of the refolding rate constant well below T_m^* indicates that the refolding process is nucleation controlled.

In our study, the Arrhenius plot of both the unfolding and folding processes showed linear dependence on temperature throughout the transition zone. As suggested by the energy profile or slower kinetics in Substitution 1, the folding process in our lattice protein was also nucleation controlled, but the maximum of the folding rate constant on the lower temperature side could not be observed. Therefore, it is likely that the absence of the maximum of the folding rate constant in our study was due to the neglect of the hydrophobic interactions.

Recently, one of the authors (F.K.)[34] investigated the folding and unfolding transition of proteins in cases with and without hydrophobic interactions by introducing a very simplified theoretical model. He showed that hydrophobic interactions of the form $f(T) = a + bT + cT^2$, $c < 0$, where T is the temperature, can account for the temperature dependence of both the specific heat and the Arrhenius plot reasonably. Hence, the expression of this type of hydrophobic interactions appears reasonable when hydrophobic interactions are explicitly taken into account in our lattice model of protein. However, his

results showed that the Arrhenius plot for the folding process is convex even in the case without hydrophobic interactions. It differs clearly from the present result for the Arrhenius plot as shown in Fig. 17.

CONCLUSIONS

Protein folding and unfolding was studied by the method of computer simulation by using a lattice model of the protein in which hydrophobic interactions are explicitly neglected. From the records of the simulation, both thermodynamic and kinetic quantities were obtained in order to discuss their temperature dependence in the transition zone and characteristics of the conformational transition of proteins. Two amino acid substituted models differing in the location of the substitution were investigated and compared with the original one to clarify the effect of the substitution on the conformational transition of the protein.

Before summarizing the results in this study, we briefly summarize the previous results obtained by the same theoretical framework as in this paper.[10-15]

(1) Specificity of the long-range interactions is important, as is non-one-dimensionality, for all-or-none character of protein folding and unfolding.

(2) The nonspecific component of the long-range interactions (a) reduces the cooperativity of the transition and (b) reduces the rate of the transitions.

(3) The short-range interactions (a) reduce the cooperativity of the transitions, but (b) accelerate the rate of the transitions.

(4) The medium-range interactions can offset the kinetic-slowing effect of the nonspecific component of the long-range interactions.

In this study, the following conclusions were reached within the limited treatment that assumed no hydrophobic interactions:

(a) The folding and unfolding transition is, to a first approximation, of all-or-none type both thermodynamically and kinetically.

(b) The difference of free energy between the N and D states, ΔG, shows linear dependence on temperature, leading to constant values of ΔH and ΔS in the transition zone.

(c) The relaxation time of the slow mode, τ_s, which reflects the overall folding and unfolding processes, shows a peak near the transition temperature T_m^*, while that of the fast mode is almost independent of temperature. The peak of the slow mode occurs at a slightly lower temperature than T_m^*.

(d) Arrhenius plots for both folding and unfolding show linear dependence on temperature. Consequently, the plot of the free energy of activation $G^{\ddagger}_{f(uf)}$ vs T^* also shows linear dependence.

(e) The values of the kinetic parameters obtained indicate that in the activated state the interactions between the units are half absent, while the state is close to the N state on the entropy axis.

(f) The amino acid substitution, which is modeled as having slightly unfavorable short-range interactions, causes the substituted proteins to be slightly unstable. Moreover, it slows the folding transition.

(g) From the analysis of the way slowing down is observed in the two substituted models, it is concluded that a structure designed to model a

β-sheet is formed before it gets assembled with other structures, which are designed to model α-helices. The process of assembly occurs nearly at the activated state of the transition.

(h) It is suggested from this study that the maximum of the folding rate constant in the Arrhenius plot, which has been observed experimentally in real proteins, is likely due to hydrophobic interactions.

APPENDIX

On the Peak of the Slow Relaxation Time

Here we derive a relation between the temperature $\hat{T}^*$, where the slow relaxation time τ_s shows a peak, and the transition temperature T_m^*. From Eq. (25),

$$1/\tau_s = k_{uf} + k_f \tag{A1}$$

We assume that the Arrhenius plot shows linear dependence on temperature as follows:

$$\ln k_f = a/T^* + c \tag{A2}$$

$$\ln k_{uf} = -b/T^* + d \tag{A3}$$

where a, b, c, and d are constants, and $a > 0$, $b > 0$. At $T^* = T_m^*$, the relation $k_f = k_{uf}$ must hold. Hence,

$$d = (a + b)/T_m^* + c \tag{A4}$$

The temperature $\hat{T}^*$ is determined by

$$d(1/\tau_s)/dT^* = 0 \tag{A5}$$

By substituting Eqs. (A1)–(A4) into (A5), we get the following relation:

$$1/\hat{T}^* = 1/T_m^* + [\ln(b/a)]/(a + b) \tag{A6}$$

Hence,

$$\text{if } b/a \geqq 1, \quad \text{then } \hat{T}^* \leqq T_m^*$$

$$\text{if } b/a < 1, \quad \text{then } \hat{T}^* > T_m^*$$

The first is the case observed in this paper. Note that Eq. (A6) is a general expression because, in the vicinity of T_m^*, the Arrhenius plot is approximately linear even in real proteins as well as this lattice protein.

We thank Drs. A. Ménez and F. M. Pohl for their valuable correspondence. Computations were carried out at the Computer Centers of Kyushu University and of the Institute for Molecular Science.

References

1. Tanford, C. (1968) *Adv. Protein Chem.* **23**, 121–282.
2. Tanford, C. (1970) *Adv. Protein Chem.* **24**, 1–95.
3. Creighton, T. E. (1978) *Prog. Biophys. Mol. Biol.* **33**, 231–297.
4. Gō, N. (1983) *Ann. Rev. Biophys. Bioeng.* **12**, 183–210.
5. Taketomi, H., Ueda, Y. & Gō, N. (1975) *Int. J. Peptide Protein Res.* **7**, 445–459.
6. Ueda, Y., Taketomi, H. & Gō, N. (1978) *Biopolymers* **17**, 1531–1548.
7. Gō, N. & Taketomi, H. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 559–563.
8. Gō, N. & Taketomi, H. (1979) *Int. J. Pept. Protein Res.* **13**, 235–252.
9. Gō, N. & Taketomi, H. (1979) *Int. J. Pept. Protein Res.* **13**, 447–461.
10. Gō, N., Abe, H., Mizuno, H. & Taketomi, H. (1980) in *Protein Folding*, Jaenicke, R., Ed., Elsevier/North-Holland Biomedical Press, Amsterdam, pp. 167–181.
11. Kauzmann, W. (1959) *Adv. Protein Chem.* **14**, 1–63.
12. Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. (1953) *J. Chem. Phys.* **21**, 1087–1092.
13. Kanô, F. & Gō, N. (1982) *Biopolymers* **21**, 565–581.
14. Glasstone, S., Laidler, K. J. & Eyring, H. (1941) *The Theory of Rate Process*, McGraw-Hill, New York.
15. Yutani, K., Ogasahara, K., Sugino, Y. & Matsushiro, A. (1977) *Nature* **267**, 274–275.
16. Yutani, K., Ogasahara, K., Kimura, A. & Sugino, Y. (1982), *J. Mol. Biol.* **160**, 387–390.
17. Privalov, P. L. & Khechinashvili, N. N. (1974) *J. Mol. Biol.* **86**, 665–684.
18. Privalov, P. L. (1979) *Adv. Protein Chem.* **33**, 167–241.
19. Schellman, J. A., Lindorfer, M., Hawkes, R. & Grutter, M. (1981) *Biopolymers* **20**, 1989–1999.
20. Hawkes, R., Grutter, M. G., & Schellman, J. (1984) *J. Mol. Biol.* **175**, 195–212.
21. Gō, N. (1975) *Int. J. Pept. Protein Res.* **7**, 313–323.
22. Oyanagi, Y. & Nakagawa, T. (1981) *SALS User's Manual* (V2.5), Tokyo University Computer Center.
23. Ménez, A., Bouet, F., Guschlbauer, W. & Fromageot, P. (1980) *Biochemistry* **19**, 4166–4172.
24. Segawa, S. & Kawai, T. (1986) *Biopolymers* **25**, 1815–1835.
25. Brandts, J. F. (1964) *J. Am. Chem. Soc.* **86**, 4291–4301.
26. Brandts, J. F. (1964) *J. Am. Chem. Soc.* **86**, 4302–4314.
27. Brandts, J. F. & Hunt, L. (1967) *J. Am. Chem. Soc.* **89**, 4826–4838.
28. Segawa, S. & Sugihara, M. (1984) *Biopolymers* **23**, 2473–2488.
29. Segawa, S. & Sugihara, M. (1984) *Biopolymers* **23**, 2489–2498.
30. Kato, S., Okamura, M., Shimamoto, N. & Utiyama, H. (1981) *Biochemistry* **20**, 1080–1085.
31. Kato, S., Shimamoto, N. & Utiyama, H. (1982) *Biochemistry* **21**, 38–43.
32. Lumry, R. & Biltonen, R. (1968) in *Structure and Stability of Biological Macromolecules*, Timasheff, S. N. & Fasman, G. D., Eds., Marcel-Dekker, New York, pp. 65–212.
33. Segawa, S., Husimi, Y. & Wada, A. (1973) *Biopolymers* **12**, 2521–2537.
34. Kanô, F. (1985) *J. Phys. Soc. Jpn.* **54**, 2347–2358.

Received July 30, 1987
Accepted October 12, 1987