Proc. Natl. Acad. Sci. USA
Vol. 75, No. 2, pp. 559–563, February 1978
Chemistry

Respective roles of short- and long-range interactions in protein folding

(mechanism of folding/range of interactions/lattice model/Monte Carlo method)

NOBUHIRO GŌ† AND HIROSHI TAKETOMI‡

† Department of Physics, Faculty of Science and ‡Computer Center, Kyushu University, Fukuoka 812, Japan

Communicated by Harold A. Scheraga, November 21, 1977

ABSTRACT **A lattice model of protein is studied by a Monte Carlo simulation method. The native conformation of the lattice protein molecule is stabilized by specific long-range and short-range interactions. By comparing results of simulation for different relative weights of the long- and short-range interactions, it is concluded that the specific long-range interactions are essential for highly cooperative stabilization of the native conformation and that the short-range interactions accelerate the folding and unfolding transitions.**

The importance of both the short-range and long-range interactions in protein folding has long been recognized. The importance of the short-range interactions was inferred first by the fair success of predicting secondary structures in the native structure of proteins from their amino acid sequences (1). The importance of the long-range interactions can be deduced from various facts, among which we will cite the following two. (*i*) Large protein fragments do not conserve the same conformations they possess in the native structure of an intact (uncleaved) protein when isolated from their complementing fragments (2), or the probability of assuming such a conformation is very low (3). This means that, for maintaining the native structure of a protein, an indispensable role is played by interfragment interactions, most of which are long range. (*ii*) Denaturational transitions in globular proteins take place in a more-or-less all-or-none manner. Even when the existence of intermediate states is discussed, the transition is certainly not of such a diffuse type as observed in the helix-coil transition of homopolypeptides. If the long-range interactions could be neglected and only the short-range interactions were assumed to be operative, then a protein could be regarded essentially as a one-dimensional system. Transitions in any one-dimensional systems are inevitably of a diffuse type (4). Therefore, the more-or-less all-or-none character of denaturational transitions in globular proteins means that the long-range interactions play an essential role in the transitions (5).

Because both the short-range and long-range interactions have been shown to be important, it is necessary to understand the respective roles of these two types of interactions in protein folding. A powerful method is to study this problem in terms of a simplified theoretical model introduced by focusing only on this point. For this purpose we incorporate short-range as well as long-range interactions into the lattice model of protein that we previously studied (6). We will describe in the present paper results obtained in the two-dimensional square lattice. A "two-dimensional protein" is admittedly a very idealized model. However, the results obtained in this paper regarding the respective roles of the short-range and long-range interactions are expected to hold in real three-dimensional proteins.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

LATTICE PROTEIN

We first consider the two-dimensional square lattice in a computer. A "protein molecule" is a self-avoiding chain polymer consisting of N units connected linearly by bonds whose length is the same as the lattice constant of the square lattice. Units are located on lattice points in the square lattice. The protein molecule assumes a variety of conformations depending on a bond angle at each unit. Two types of forces are assumed to be operative in this molecule, the long-range and short-range interactions.

The long-range interactions are assumed to work in such a way that the energy of the system decreases by ϵ when one of preassigned pairs of units occupies the nearest-neighbor lattice points. The assignment of pairs of interactable units defines the specificity of the long-range interactions. We consider a "polymer chain" consisting of 49 units linked linearly. In the previous paper (6) we assumed three different specificities, strong limit specificity A, intermediate specificity B, and weak limit specificity (i.e., no specificity) C. In specificity A, only 36 pairs of units, occurring in the nearest-neighbor lattice points in the conformation in Fig. 1, are assigned. These 36 pairs are indicated by black squares in Fig. 2. (For example, unit 1 has three units, 4, 6, and 8, as its nearest neighbor in Fig. 1. Therefore, pairs 1-4, 1-6, and 1-8 are assigned to be interactable in specificity A and are so indicated in Fig. 2 by black squares.) Therefore, the conformation in Fig. 1 has the long-range interaction energy of -36ϵ. However, any other conformations, grossly different from the one in Fig. 1, have the long-range interaction energy much closer to zero, even when such conformations are compactly packed. This is because a pair assigned in specificity A occupies nearest-neighbor lattice points in such conformations only with very low probabilities. The conformation shown in Fig. 1 has an extraordinarily low interaction energy. Therefore, for specificity A the "polymer chain" is expected to have the conformation in Fig. 1 as its native structure at low temperatures. In specificity C, all 552 pairs of units that can geometrically occupy the nearest-neighbor lattice points are assigned. In this case all conformations compactly folded into a 7×7 square have the same lowest energy, -36ϵ. Therefore, no particular conformation will be chosen as a native structure. In specificity B, randomly selected 148 pairs, as well as 36 pairs that are assigned in specificity A, are assigned. These 148 pairs are indicated by shaded squares in Fig. 2. As a result 184 pairs are assigned. (For example, unit 1 is assumed to be interactable with units 4, 6, 8, 14, 22, and 28.) This is exactly one-third of the 552 pairs assigned in specificity C. Therefore, most of the conformations compactly folded into a 7×7 square are expected to have the interaction energy of about -12ϵ (i.e., $\frac{1}{3}$ of -36ϵ). However, the conformation in Fig. 1 still has an extraordinarily low interaction energy of -36ϵ because all pairs assigned in specificity A are retained in specificity B. Therefore, the conformation in Fig. 1 is also

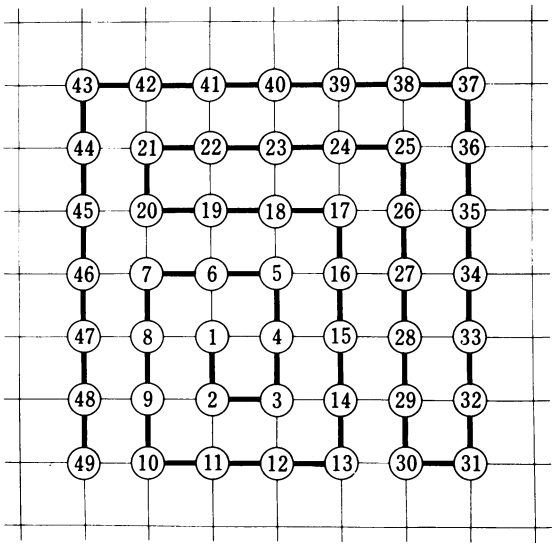

FIG. 1. Native conformation of a "protein" in the two-dimensional square lattice.

expected to be the native structure at low temperatures for specificity B.

In the previous paper conformational changes of this lattice protein were studied by the Monte Carlo method of Metropolis *et al.* (7). Only the long-range interactions with the above three different specificities were assumed to be operative. For specificity A, reversible all-or-none type transition between the native and denatured state was observed. For specificity C, compact (or globular) but not specific conformations are stable at low temperatures. The globule-coil transition is reversible but of the diffuse type, as in the helix-coil transition. Therefore, the specificity of the long-range interactions is essential for the all-or-none character of the transition. For specificity B, regarded to be the most realistic of the three, the conformation of Fig. 1 is stable at low temperatures and a transition into the denatured conformation occurs quickly, suggesting an all-or-none type transition. However, the transition was not reversible within the time range of the computer experiment. Therefore, in the present paper we take specificity B as the long-range interactions, and we will study effects of inclusion of the short-range interactions as described below.

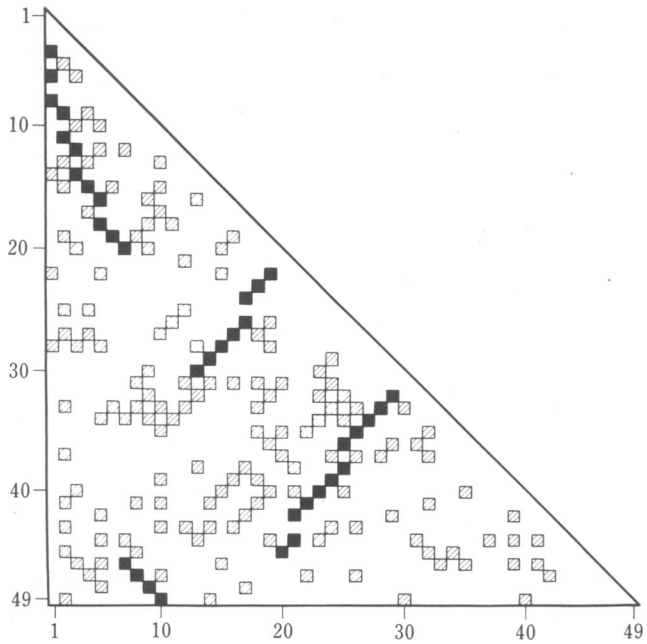

FIG. 2. Specification of interactable pairs in specificity A (black squares) and in specificity B (black and shaded squares). Both abscissa and ordinate indicate residue number.

As the short-range interactions, we consider such an energy term that is a function of a "bond angle" at each unit. A bond angle at the ith unit is defined as the angle between two bond vectors, one pointing from $(i-1)$th unit to the ith unit, and the other from the ith unit to the $(i+1)$th unit. The bond angle at each unit can assume three different values, 90°, 180°, and 270°. As the short-range interactions, we assume a certain value of the "bond energy" for each of these three values of each bond angle. In this paper we assume that, when a bond angle at ith unit takes the same value as in the conformation in Fig. 1, the bond energy of the ith unit is lower by ϵ' than two other possible cases of the value of the angle ($i = 2,3,\cdots,48$). This is a good model of the short-range interactions in real proteins in the sense that they determine conformational propensity of individual units or residues. However, the assumption of the short-range interactions, whose energy is the lowest at the native conformations, is an idealization. The short-range interactions in real proteins, though known to be important, are not perfectly consistent with native conformations because otherwise it would be possible to predict native conformations from a knowledge of short-range interactions only.

The Monte Carlo method of Metropolis *et al.* (7) is again used in this paper. Computer experiments have been performed by taking four different ratios of the values of ϵ and ϵ', i.e., $(\epsilon,\epsilon') = (0.75\epsilon_0,0.25\epsilon_0)$, $(0.5\epsilon_0,0.5\epsilon_0)$, $(0.25\epsilon_0,0.75\epsilon_0)$, and $(0,\epsilon_0)$. Here, ϵ_0 is the unit of interaction energy, which will be used to define the dimensionless temperature $T^* = kT/\epsilon_0$, where k is the Boltzmann constant. The case of $(\epsilon,\epsilon') = (\epsilon_0,0)$ is the same as the case of specificity B studied in the previous paper (6). By comparing the results obtained for these cases, we can infer relative roles of the short-range and long-range interactions in protein folding and unfolding.

RESULTS

In each case of the ratio of ϵ and ϵ', computer experiments of conformational changes were carried out at a series of values of temperature T^* mainly in order to determine the transition temperature T^*_m. At each temperature two different initial conformations were chosen, the native conformation of Fig. 1 and a conformation in which 49 units are on a straight line. The latter was chosen as an example of a denatured conformation. Once the transition temperature T^*_m is determined, very long computer experiments were carried out at $T^* = T^*_m$. Detailed analysis of the results will be published elsewhere. In this paper only a few records of experiments are shown which are relevant for the discussion of respective roles of the long- and short-range interactions.

Figs. 3–6 show records of computer experiments for the four cases of (ϵ,ϵ'). The trial number on the abscissa can be regarded approximately proportional to physical time (6). The ordinate indicates the degree of the order of the system, represented by the interaction energy counted in the unit of $(-\epsilon_0)$. The broken lines indicate the expected count m of the short-range interaction energies at the high temperature limit. At any finite temperatures average values of the count of the short-range interaction energies are higher than these values. We may regard only the ranges of the energy count m between the broken lines and the upper limit values as physically meaningful.

At low temperature $T^* = 0.3$, folding into the native con-

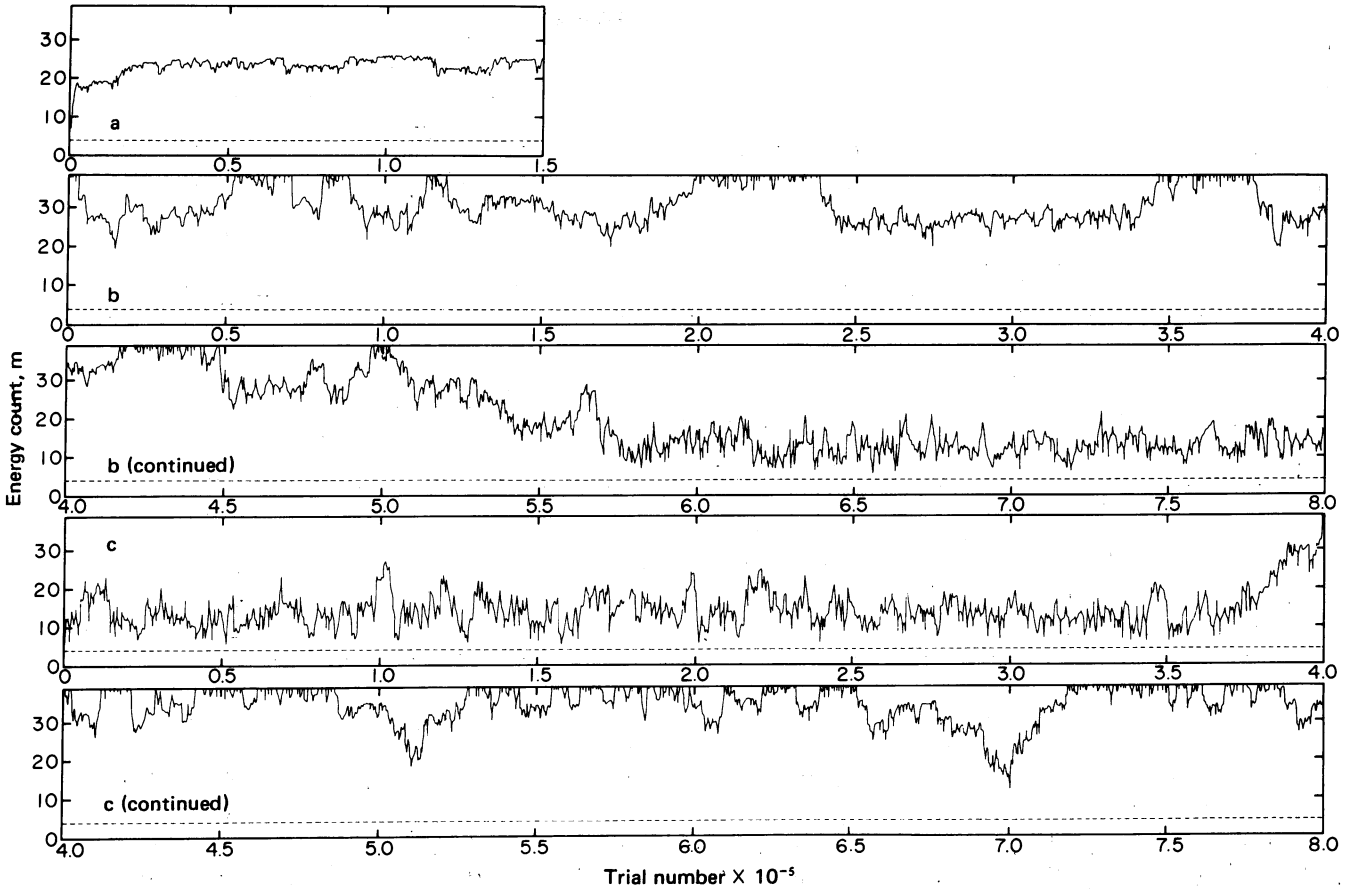

FIG. 3. Records of conformational changes in the case of $(\epsilon,\epsilon') = (0.75\epsilon_0, 0.25\epsilon_0)$. The ordinate is the conformational energy counted in the unit of $(-\epsilon_0)$. The broken line at $m = 0.25 \times 47 \times (1/3)$ indicates the average short-range interaction energies at the high temperature limit, where each of 47 bond angles assumes three possible values equally probable, one of which has an energy of $\epsilon' = -0.25\epsilon_0$. Only the range between the broken line and the upper limit may be regarded as physically meaningful. (*a*) $T^* = 0.3$, starting from a denatured conformation. (*b* and *c*) $T^* = 0.675$, starting from the native and a denatured conformation, respectively.

formation is not observed within 1.5×10^5 trials in Fig. 3*a*. The lattice polymer is trapped in a local minimum with about $m = 25$ and cannot get out of it. In Fig. 4*a*, folding into the native conformation is observed at about the 0.6×10^5th trial. In Fig. 5*a* folding into the native conformation occurs quickly within 0.2×10^5 trials. It becomes very quick in Fig. 6*a*.

At temperatures close to the transition temperatures T_m^*, both folding and unfolding transitions are observed in Fig. 3*b* and

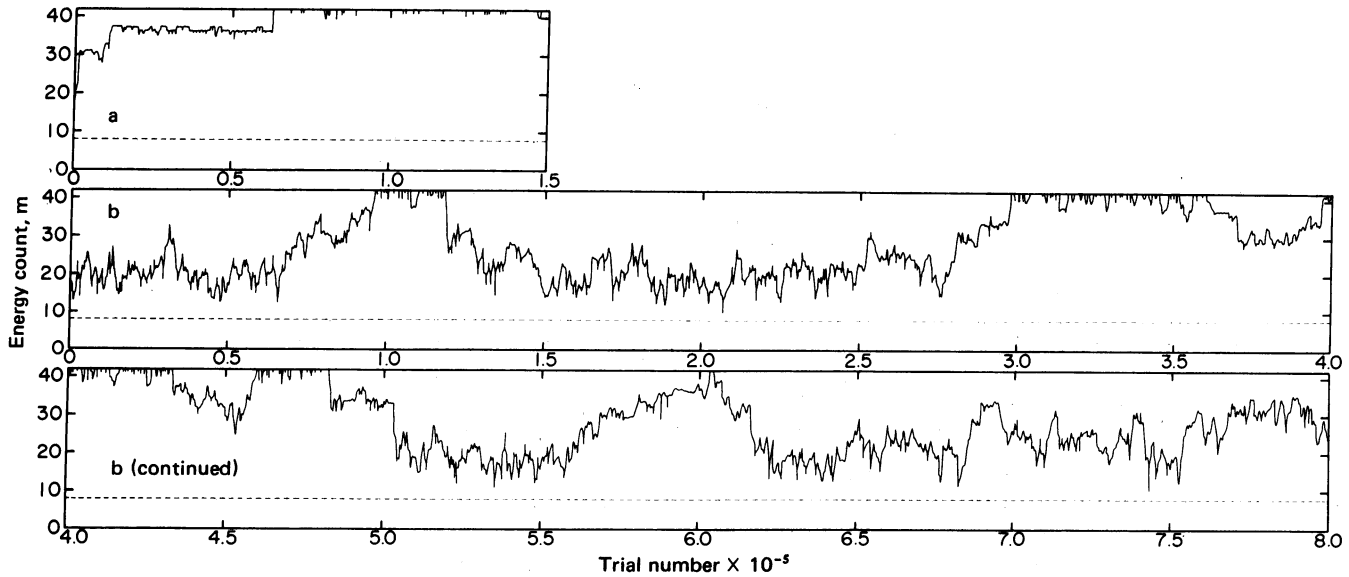

FIG. 4. Same as in Fig. 3, but for $(\epsilon,\epsilon') = (0.5\epsilon_0, 0.5\epsilon_0)$. (*a*) $T^* = 0.3$ and (*b*) $T^* = 0.6$, both starting from a denatured conformation.

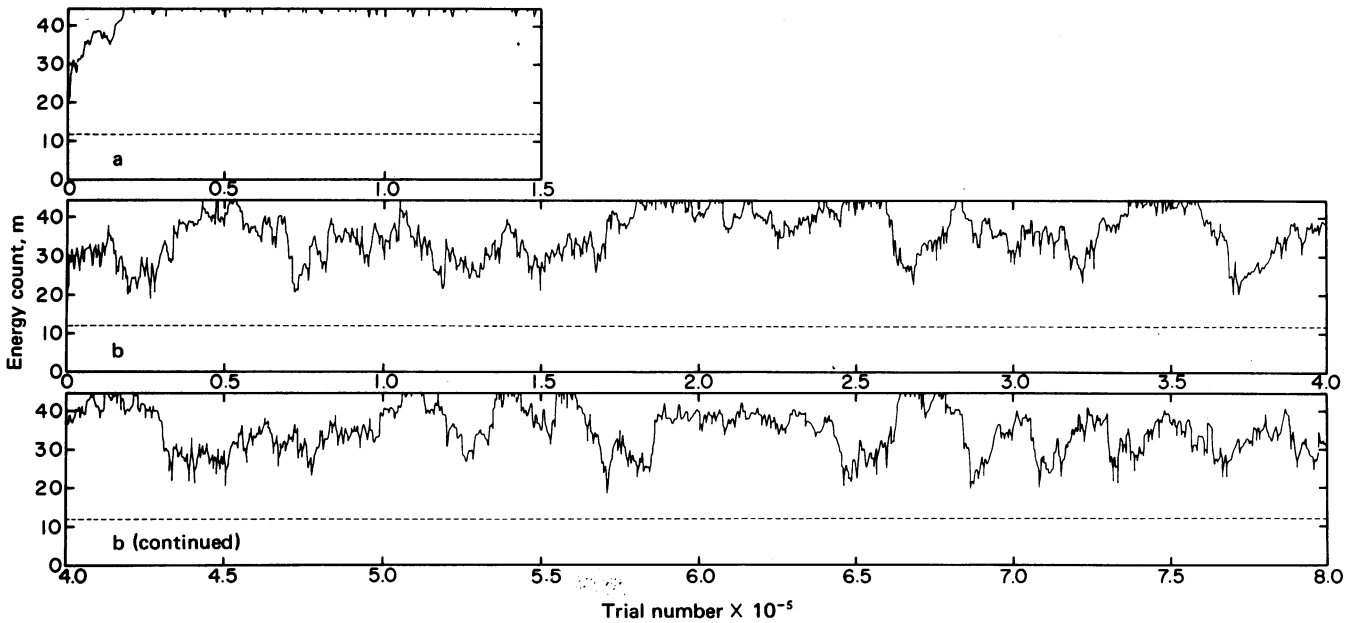

FIG. 5. Same as in Fig. 3, but for $(\epsilon,\epsilon') = (0.25\epsilon_0, 0.75\epsilon_0)$. (*a*) $T^* = 0.3$ and (*b*) $T^* = 0.525$, both starting from a denatured conformation.

c. However, the transitions are very rare, which indicates the strong all-or-none character of the transition. The transitions occur more frequently in Fig. 4*b*. In Fig. 5*b*, the transitions become so frequent that populations at intermediate states are as significant as those of the completely folded and unfolded conformations. In Fig. 6 *b* and *c*, where there are no long-range interactions except for the self-avoidance of the chain, the lattice protein is fluctuating around the average value of about $m = 34$, i.e., there is only one peak of population centering at $m = 34$. This indicates that the folding and unfolding transitions are of the diffuse type, as in the helix-coil transitions.

The most striking fact shown in Figs. 3–6 is that folding into the native conformation is possible while the folding could not be observed when only the long-range interactions with specificity B were assumed. The effect of the short-range interactions, to accelerate the folding process, is obvious. By comparing Figs. 3–6 it is clearly seen that the folding is more accelerated as the contribution from the short-range interactions is more weighted.

The short-range interactions have another effect, that of making the transition diffuse. This effect is clearly seen by comparing Fig. 3 *b* and *c*, Fig. 4*b*, Fig. 5*b*, and Fig. 6 *b* and *c*. In the extreme case of the short-range interactions only, the transition is no longer of the type of the denaturational transitions in globular proteins. It is concluded that the specific long-range interactions are essential for causing the transition to be of the all-or-none type or, in other words, for rendering the native conformation highly cooperatively stabilized.

DISCUSSION

The observed acceleration of the folding and unfolding transitions by the short-range interactions may be explained by two different mechanisms. As the all-or-none character of the transition becomes less pronounced as the short-range interactions are more weighted, the free energy of activation for the folding and unfolding transitions becomes lower, i.e., the probability of the least probable state between folded and unfolded states becomes high. This renders the transition kinetically faster.

Another possible mechanism is the following. The short-range interactions promote short-range conformational order, i.e., short segments of the chain have higher probabilities of assuming the same local conformations as in the native conformation. These local "correct" conformations, though unstable

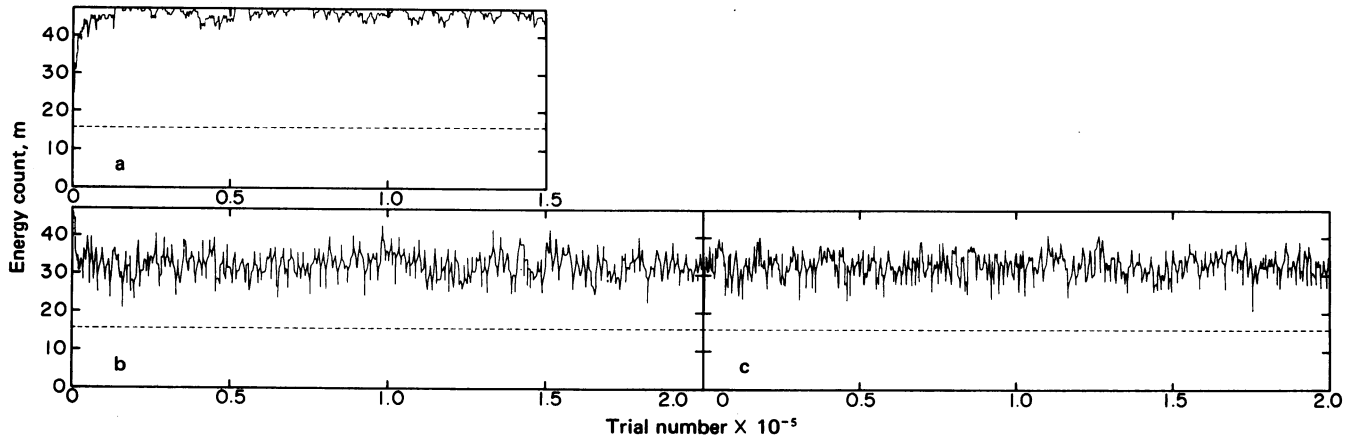

FIG. 6. Same as in Fig. 3, but for $(\epsilon,\epsilon') = (0,\epsilon_0)$. (*a*) $T^* = 0.3$, starting from a denatured conformation. (*b* and *c*) $T^* = 0.7$, starting from a native and the denatured conformation, respectively.

by themselves, may serve as structural units for making larger structures. Folding of these local structural units into the native conformation should be much faster than the case in which there are no significant local structures and all units in the chain tend to behave individually. Both of these two mechanisms may be necessary to explain the observed acceleration.

These two mechanisms of the acceleration of the folding and unfolding transitions do not depend much on details of long- and short-range interactions, but depend essentially only on the ranges of interactions, i.e., some are long and the others are short. Therefore, these two mechanisms are expected to exist also in real globular proteins. The design and analysis of an experiment to measure these two mechanisms more quantitatively and to see how these two are combined is a problem raised here for future studies.

The computation was done at the Computer Center, Kyushu University. This work was supported by grants from the Ministry of Education, Japan.

1. Scheraga, H. A. (1973) *Pure Appl. Chem.* **36**, 1–8.
2. Epand, R. M. & Scheraga, H. A. (1968) *Biochemistry* **7**, 2864–2872.
3. Anfinsen, C. B. (1972) *Biochem. J.* **128**, 737–749.
4. Landau, L. D. & Lifshitz, E. M. (1958) in *Statistical Physics* (Pergamon Press, London), p. 482.
5. Gō, N. (1976) in *Advances in Biophysics*, ed. Kotani, M. (Tokyo University Press, Tokyo), Vol. 9, pp. 65–113.
6. Taketomi, H., Ueda, Y. & Gō, N. (1975) *Int. J. Protein Peptide Res.* **7**, 445–459.
7. Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. (1953) *J. Chem. Phys.* **21**, 1087–1092.