Int. J. Peptide Protein Res. 7, 1975, 445–459
Published by Munksgaard, Copenhagen, Denmark

STUDIES ON PROTEIN FOLDING, UNFOLDING AND FLUCTUATIONS BY COMPUTER SIMULATION

I. The effect of specific amino acid sequence represented by specific inter-unit interactions

HIROSHI TAKETOMI, YUZO UEDA and NOBUHIRO GŌ

Department of Physics, Faculty of Science, Kyushu University, Fukuoka, Japan

Received 9 December 1974

A lattice model of proteins is introduced. "A protein molecule" is a chain of non-intersecting units of a given length on the two-dimensional square lattice. The copolymeric character of protein molecules is incorporated into the model in the form of specificities of inter-unit interactions. This model proved most effective for studying the statistical mechanical characteristics of protein folding, unfolding and fluctuations. The specificities of inter-unit interactions are shown to be the primary factors responsible for the all-or-none type transition from native to denatured states of globular proteins. The model has been studied by the Monte Carlo method of Metropolis et al., which is now shown applied to approximately simulating a kinetic process. In the strong limit of the specificity of the inter-unit interaction the native conformation was reached in this method by starting from an extended conformation. The possible generalization and application of this method for finding the native conformation of proteins from their amino acid sequence are discussed.

This is the first paper in a series devoted to the study of the conformational properties of proteins from the statistical mechanical point of view. In this series we focus attention specifically on conformational fluctuations of globular proteins in the native state and on the processes of folding and unfolding. We would like to justify such a statistical mechanical study by two significant possibilities. Firstly, conformational fluctuations in the native state are expected to be important for the understanding of the functions of proteins. Secondly, it may help to find a way to develop an algorithm for predicting the three-dimensional structure of proteins from their amino acid sequence, this is a problem of the most fundamental biological significance and a theoretical challenge. The basic strategy tried so far has been to minimize the conformational energy and to locate the global minimum as the native conformation. However, this strategy did not work because of the enormous number of local minima. Because of this we decided to study the real process of folding from denatured to native conformation. Folding and unfolding are processes of a statistical mechanical nature.

From the viewpoint of statistical mechanics, proteins are three-dimensional, inhomogeneous (or non-repetitive) finite chain systems. The specificity of the native conformation of proteins, or their ability to assume unique conformations in their native states, is a striking aspect of such systems. It was shown in the previous paper (1) that the globularity and specificity of the native conformations of proteins were the essential factors determining the statistical mechanical properties of the conformations of proteins (e.g. the fluctuations and transitions). The specificity of the native conformations of proteins was treated as a given phenomenological fact and its influence on the character of the conformational

transition was discussed. In this paper the specificity is treated, not as a given fact, but as deriving from the heterogeneity of the amino acid sequence of a protein.

In order to study the statistical mechanical properties of denaturation and conformational fluctuations of proteins and to understand the mechanism of realization of specificity of native conformations of proteins, we introduce a lattice model of proteins into which heterogeneous aspects in the amino acid sequence are incorporated. In statistical mechanical treatments of polymers in general, inter-unit interactions are conceptually divided into short-range and long-range interactions. Interactions between units that are separated far along the chain are defined to be long-range, even when the units are not widely separated in space. This division of the interactions into short-range and long-range ones is important also in theoretical treatments of protein conformations. Existence of the correlations between short-range amino acid sequence and secondary structures (e.g., α-helix, β-structure and β-turn) in globular proteins has been shown recently by many authors. This indicates that the short-range interactions have considerable importance in the process of folding, even though it is apparent that the long-range interactions have their own role in determining and stabilizing the native conformation of proteins. This role of the long-range interaction has not yet been sufficiently studied theoretically because of the theoretical difficulties intrinsic in the long-range interactions. Differences in amino acids manifest themselves in heterogeneities in both the short-range and long-range interactions. In this paper emphasis is placed on the role of the long-range interactions. The short-range interactions are not discussed. The heterogeneity of the long range interactions is taken into the lattice model by a very idealized form of the specificity of inter-unit interactions, i.e. attractive interactions are assumed to work for a set of specifically preassigned pairs of units occupying the nearest neighbor lattice points. In section I. the model is described. Even though abstracted and simplified, the model is still far from being amenable to analytical treatment. In this paper the model is studied by the Monte Carlo simulation method of Metropolis et al. (2). The method of Monte Carlo calculation is described in section II. In section III, the results of the simulations are presented and discussed. The purpose of abstracting protein molecules as a lattice polymer is to enable otherwise complex phenomena to be studied in as pure and simple a form as possible. Because of the simplification and idealization, some of the conformational properties of proteins become lucidly understandable. However, this is achieved by discarding some of the important aspects of real proteins. The discussion of these aspects is also included in section III. In section IV brief conclusions and the outlook for further studies are given.

I. DESCRIPTION OF THE MODEL

In this paper we are considering a lattice polymer on the two-dimensional square lattice. A study on a lattice polymer on the three-dimensional cubic lattice is now in process and will be presented in a future paper. The polymeric chain on the lattice is assumed to be self-avoiding due to repulsive interactions between two units within the contact distance. The heterogeneity in the amino acid sequence is incorporated as the specificity of inter-unit interactions. Attractive interactions are assumed to work only for a set of preassigned pairs of units occupying the nearest neighbor lattice points. Such pairs of units are defined as being interactable and are indicated by filled (black) squares in Fig. 1. A pair of units remains unfilled (white), i.e. the pair is not interactable, when no interaction works between them even if the pair occupies the nearest neighbor lattice points. Thus Figs. 1A, B and C represent the inter-unit interactions with the three different types of specificity, respectively.

The distribution of interactable pairs in Fig. 1A is determined as follows. First the native conformation of a protein consisting of 49 units is assumed on the square lattice as in Fig. 2. All pairs of units occupying the nearest neighbor lattice points in the conformation of Fig. 2 are filled black and all other pairs are unfilled in Fig. 1A. The energy of inter-unit interaction is assumed to be identical for all interacting pairs with a value of $-\varepsilon$ in this paper. Therefore the conformational energy of this protein molecule is measured in the unit of $(-\varepsilon)$. Since all neighboring pairs in Fig. 2 are interacting for specificity A specified by Fig. 1A, the conformational energy of the conformation of Fig. 2 is -36ε for

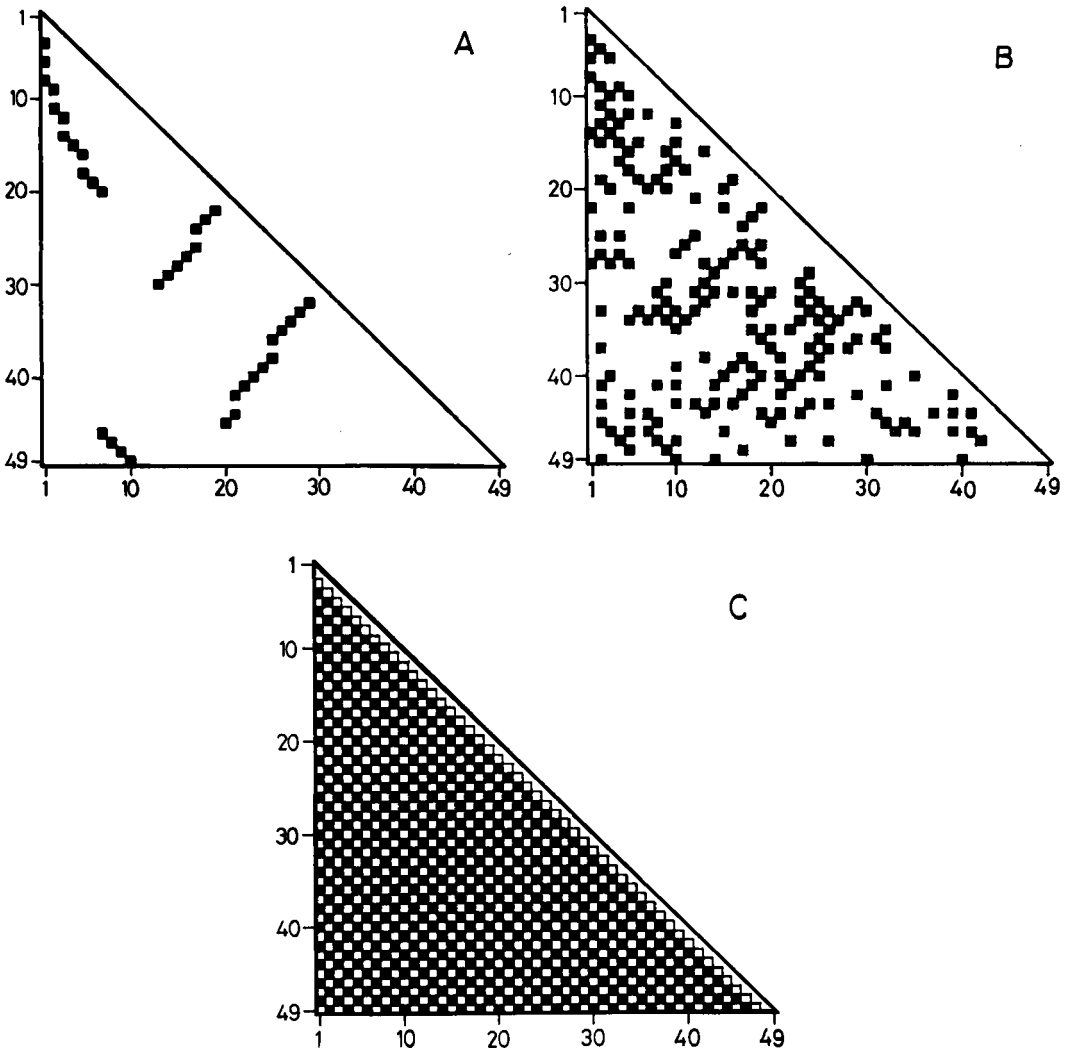

FIGURE 1

Specificity of inter-unit interactions. Attractive interaction is assumed to work when a filled (black or shaded) pair of units occupy the nearest neighbor lattice points. A: strong limit specificity, B: intermediate specificity, C: weak limit specificity.

this specificity of interaction. It is evident that any conformations other than those shown in Fig. 2 (and its mirror image) have conformational energies higher than -36ε for this specificity. Therefore for specificity A, the native conformation, i.e. the conformation with the global minimum energy, is that given in Fig. 2. (No distinction will be made between a conformation and its mirror image in this paper.)

In Fig. 1C all 552 pairs of units separated by $(2n+1)$ bonds $(n \geqslant 1)$ are filled black, i.e. attractive interaction works nonspecifically for any pair of units occupying the nearest neighbor lattice points. No other pairs of units can occupy the nearest neighbor lattice points due to geometrical reasons. Because no specificity exists, the case of Fig. 1C is reduced to a homogeneous nonintersecting chain polymer with short-range

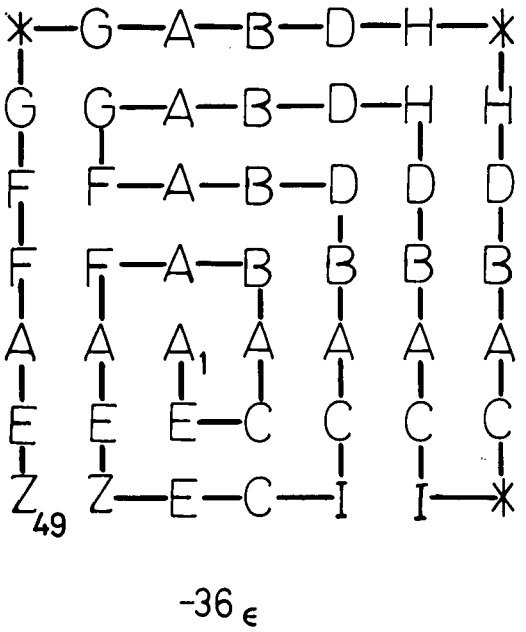

FIGURE 2
Native conformation of "a protein" on a two-dimensional lattice.

attractive interaction. There is no unique conformation with the global minimum energy. In fact, any conformations packed compactly into a 7×7 square, not only the conformation of Fig. 2 but also any of the conformations like those in Fig. 3, have the global minimum energy -36ε. For the specificity specified in Fig. 1C (i.e. no specificity), the chain polymer should have a tendency to assume compact globular forms at low temperatures, but there is no unique native conformation even at the low temperature limit.

The two cases in Figs. 1A and C can be regarded as the limits of the strongest and weakest specificities, respectively. It is possible to consider cases of intermediate specificities. Fig. 1B offers one such example. The specificity of Fig. 1B is determined as follows. All 36 interactable pairs of units in Fig. 1A remain interactable in Fig. 1B. In addition to them, randomly selected pairs of units are taken to be interactable, increasing the total number of interactable pairs of units to 184; exactly one-third of the 552 interactable pairs in the case of the weakest specificity in Fig. 1C. For this specificity B the conformational energy of the conformation in Fig. 2 is -36ε, since all interactable pairs in Fig. 1A are retained in Fig. 1B. However, other compact conformations in Fig. 3 have higher conformational energies of -10ε, -12ε and -13ε from left to right, respectively. An arbitrary compact conformation in 7×7 square is expected to have an average conformational energy of -12ε, because one-third of all pairs are assigned as interactable in this specificity. Therefore, of all the compact conformations, the one in Fig. 2 has distinctly low conformational energy, indicating that at sufficiently low temperatures, the conformation of Fig. 2 should be the unique one except in a very improbable accidental case.

The intramolecular interactions in real proteins are conceptually classified into the long-range and

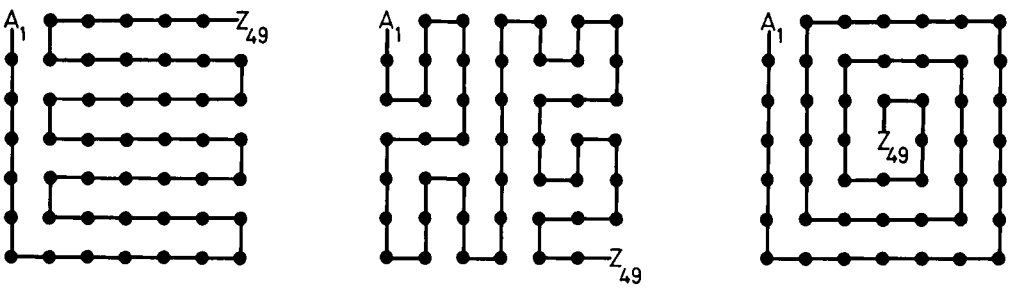

FIGURE 3
Examples of arbitrary compact conformations other than the native one. Each unit is assigned either a letter of the alphabet or a symbol (*). The same letter of the alphabet is given to a pair of units which are interactable in specificity A. This facilitates observing which pairs are interacting in an arbitrary conformation, e.g. in conformations in Fig. 12, a pair of units occupying the nearest neighbor points is not interacting when the pair of units does not bear the same letter of the alphabet.

short-range interactions as discussed in Introduction. The long-range interactions are represented in the lattice model by the inter-unit interactions so far discussed. The short-range interactions in real proteins endow each amino acid residue with a tendency to assume an intrinsically preferred backbone conformation. A possible representation in the lattice model of such short-range interactions is as follows. We define a "bond angle" of the i-th unit in the lattice model as an angle between "bonds" connecting units (i-1) and i, and i and (i+1). The bond angle assumes one of the three values 0° and ±90° in the model of a two-dimensional square lattice. We define an intrinsic bond angle of the i-th unit as the bond angle realized in the native conformation. The short-range interaction can be incorporated into the lattice model by assigning conformational energies which depend on the bond angles in such a way that the energy is lower for angles assuming the intrinsic value than for angles assuming a non-intrinsic value. In the present paper, however, we focus attention on the long-range interactions. The short-range interactions are not discussed in this paper.

The purpose of the present paper is, as described in the Introduction, to study the statistical mechanical structure of conformational transitions and fluctuations. In the cases of specificities shown in Figs. 1A and B, the native conformations are expected to be unique. The case of Fig. 1A appears to be somewhat artificial as a model of proteins, as attractive interactions are assumed only for the pairs which are interactable in the native conformation. In this respect, the case of Fig. 1B looks more natural. By studying the three cases of different specificities in Figs. 1A, B and C, we should be able to find the influences that specificities of interaction have on the statistical mechanical mechanisms of conformational transitions and fluctuations of a chain polymer which assumes a specific and globular conformation in the native state.

II. THE METHOD OF COMPUTER ANALYSIS

Because the model is not amenable to analytical treatments, the model has been studied by the Monte Carlo simulation method of Metropolis et al. (2) The method consists in constructing a Markov process of conformational transitions so that in the stationary state an individual conformational state has the probability of occurrence proportional to that determined by the Boltzmann law. The transition probability p_{ij} of the Markov process from conformational state i to conformational state j is determined by (a) the a priori transition probability p^0_{ij} from state i to state j, and (b) conformational energies E_i and E_j of state i and state j, respectively. The a priori transition probabilities satisfy the condition of detailed balance,

$$p^0_{ij} = p^0_{ji} \tag{1}$$

i.e. the a priori transition probability from state j to state i is equal to that from state i to state j. The transition probability is then given by

$$p_{ij} = \begin{cases} p^0_{ij} & \text{for } E_i - E_j \geqq 0 \\ p^0_{ij} \exp[(E_i - E_j)/kT] & \text{for } E_i - E_j < 0 \end{cases} \tag{2}$$

It is demonstrated in ref. (2) that in the stationary state of the Markov process defined by equation 2 an individual state has the probability of occurrence proportional to that determined by the Boltzmann law. The temperature comes into the model through equation 2. The temperature will be measured hereafter in the unit of ε/k and designated by T^*. The conformational energy E_i of state i is given by $-m_i\varepsilon$, where m_i is the number of interactable pairs occupying the nearest neighbor lattice points.

The Monte Carlo method of Metropolis et al. was originally devised as a modified Monte Carlo integration over configuration space so that the equilibrium values of physical quantities could be evaluated for a system undergoing phase transitions. If the transition probabilities are chosen properly, however, this method can be applied to simulate approximately the time course of a system behaving stochastically. If the rate constant for a system to change from state i to state j is k_{ij}, a stochastic model of the kinetic behavior of the system can be defined by assigning $k_{ij}\Delta t$ for a transition probability $p_{ij}(\Delta t)$ from state i to state j in time interval Δt. A stochastic model of discrete time can be defined by choosing such a small time interval Δt as the unit time length during which the transition of state takes place once at the most. In this discrete-time stochastic model the transition probability $p_{ij}(\Delta t)$ is very small, and satisfies the relation

$$p_{ij}(\Delta t)/p_{ji}(\Delta t) = \exp[(E_i - E_j)/kT]. \quad (3)$$

The probability $p_{ii}(\Delta t)$ of staying in state i during the time interval Δt is given by

$$p_{ii}(\Delta t) = 1 - \sum_{j \neq i} p_{ij}(\Delta t) \quad (4)$$

which is very close to unity, i.e. the transition is quite seldom. The course of simulation calculation can be interpreted exactly as a kinetic process. The calculation, however, requires long computing time, because transitions are quite seldom. Computing time can be reduced by taking long Δt, so that the transition probabilities proportionally become larger, although the length of the unit time Δt is limited due to the transition taking place once at the most in the interval. The transition probabilities, which still satisfy equation 3, are determined by the method of Metropolis et al. so that transitions take place reasonably fast in computer calculation. Thus, in the method a short unit time length Δt is not taken, yet transition is assumed to take place once at the most, i.e. the probabilities of more than one transition taking place in the unit time length are discounted. This is the approximation involved in the kinetic interpretation of the method of Metropolis et al. Except for this approximation, the course of simulation calculation can be interpreted kinetically, if the transition probabilities are taken realistically. For this purpose the a priori probability p^0_{ij}, which is equal to $(p_{ij}p_{ji})^{1/2} \exp(|E_i - E_j|/2kT)$, should be taken as close to $(k_{ij}k_{ji})^{1/2}\exp(|E_i - E_j|/2kT)\Delta t$ as possible. This is the reason why higher a priori probabilities are assigned in this paper to easier movements (as described in the next paragraph). These considerations led us to regard the course of simulations calculation approximately as a kinetic process; in other words, trial numbers in Figs. 6 and 11, can be regarded as the quantities equivalent to time.

The a priori transition probability p^0_{ij} is determined as follows. The a priori transition probability p^0_{ij} has a non-vanishing value only when the transition from state i to state j is attained by one of the elementary conformational changes. Fig. 4 illustrates the five types of elementary changes; (1) non-local rotation, (2) non-local translation, (3) local translation a, (4) local translation b and (5) local translation c. One unit in each is filled in Fig. 4. If this is j^{th} unit in the chain, each of the conformational changes are said to be taking place at j^{th} unit. The flow chart of the actual computational procedure is shown in Fig. 5. The conformational changes of types 1 through 5 are chosen randomly with the probabilities of 10%, 25%, 26%, 21% and 18%, respectively. A unit

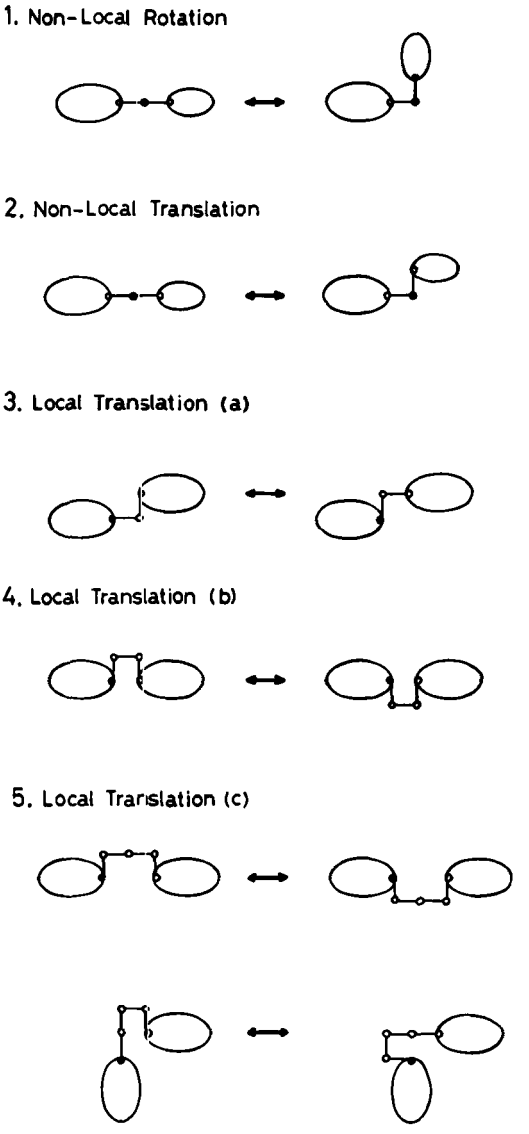

FIGURE 4

Five types of elementary conformational changes. Each of the changes takes place at the filled unit in order to distinguish from the same type of change taking place at different units in the chain.

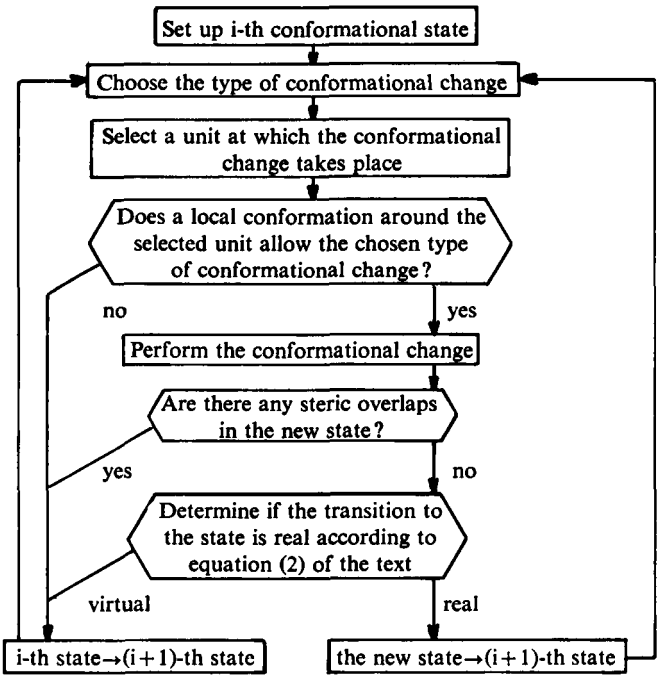

FIGURE 5
Flow chart of the actual computational procedure.

undergoing one of the three types of local conformation al changes is selected randomly with uniform probabilities from all units. For the two types of non-local conformational changes, the probability of choice is increased toward both ends of a chain showing that a change could occur more easily when an involved moving part has a smaller moment of inertia. The probability thus assigned determines the a priori probability that a conformational change of chosen type takes place at a chosen unit, provided (a) that such a conformational change can actually occur at the chosen unit [e.g. the conformational change of type 3 can not occur at a unit whose "bond angle" is 0°] and (b) that there is no steric overlap in a conformation resulting from the conformational change. It is apparent that the a priori probabilities thus determined satisfy the condition of detailed balance.

III. RESULTS AND DISCUSSION

Fig. 6 shows the time course of conformational changes starting from the native conformation of Fig. 2 in the cases of the interaction specificities A and C for various temperatures. The abscissa is the number of trials of conformational changes which can be approximately regarded as time as described in section II. The ordinate indicates the degree of the order of the system which is represented by the number m of attractive interactions in a given conformation. In the case of specificity C, after relaxation to the equilibrium state, m fluctuates around a mean value $<m>$, which decreases gradually as the temperature is raised. In the case of specificity A, the chain polymer remains fluctuating about the native state at low temperatures, with the amplitude of fluctuation getting larger as the temperature is raised to melting point. As the temperature approaches melting point, the chain polymer at first remains fluctuating about the native state for a while and then suddenly loses the order to become the denatured state. At temperatures near melting point, the once denatured polymer suffers a sudden large-scale fluctuation, and returns to the native state ($T^* = 0.85$ in Fig. 6A). As the temperature is raised beyond melting point, the once denatured polymer remains fluctuating at the denatured state ($T^* = 1.2$ in Fig. 6A). An

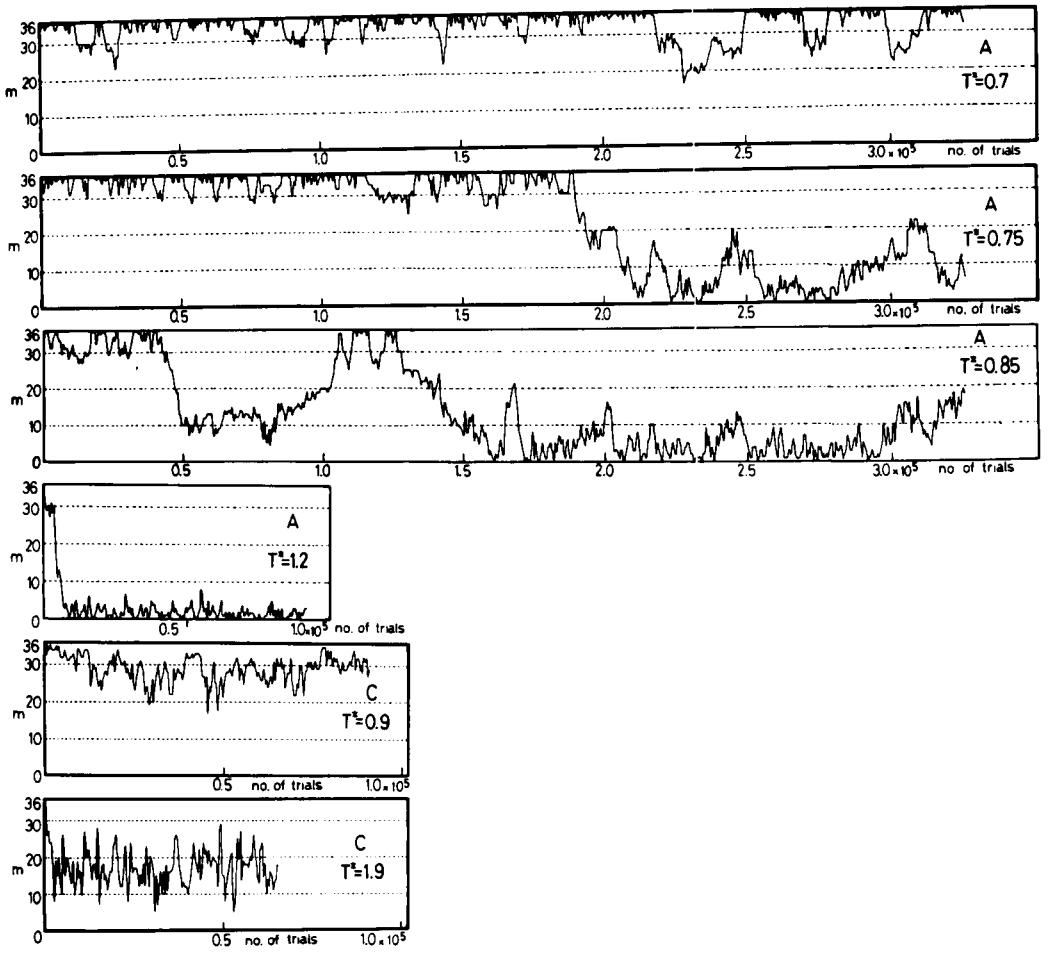

FIGURE 6

Courses of conformational changes starting from the native conformation of Fig. 2 in cases A and C at various temperatures.

effect of the interaction specificity is clearly seen from the qualitative differences of the curves for cases A and C in Fig. 6.

By averaging the curves of Fig. 6 for a large number of trials, we can get an equilibrium value of θ defined by $<m>/m_{max}$ at each temperature, where m_{max} is the maximum number of interactable pairs, i.e. $m_{max} = 36$ in the present model. The equilibrium denaturation curves (i.e. θ versus T curves) are shown in Figs. 7A, B and C for the cases of specificities A, B and C, respectively. It is clearly seen that the transition becomes sharper as the interaction becomes more specific. However, such a basic character of the transition, as to whether or not it is of the all-or-none type, cannot be deduced from the transition curves in Figs. 7A, B and C alone. In order to elucidate such a problem, the population distribution in equilibrium is plotted in Fig. 8 for two temperatures near the transition temperature in the case of specificity A. At both temperatures the distribution has two peaks, one at $m = 36$ (native state) and the other at $m \simeq 0$ (denatured state). Therefore by definition (1) the transition is of the all-or-none type. As was demonstrated in ref. (1), the character of the conformational transition can be fully described

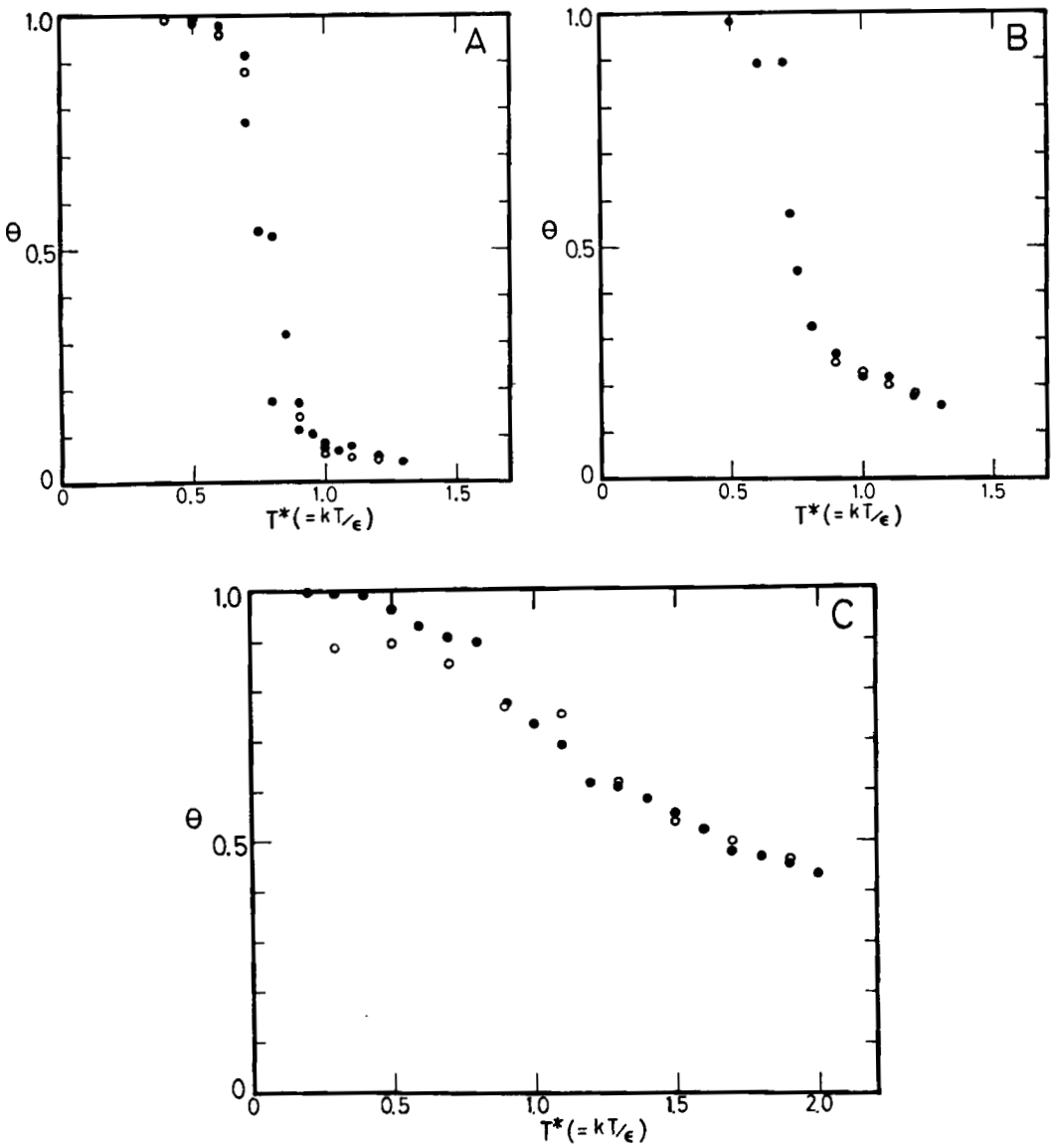

FIGURE 7

Equilibrium denaturation curves for the three cases of specificity A, B and C. ● and ○ are data obtained by starting the simulation from the native and denatured conformations, respectively.

by S-H curves, which can be constructed from population distributions like those in Fig. 8. S-H curves constructed from computed data are shown in Figs. 9A, B and C for the three cases of specificities A, B and C, respectively. The curve of Fig. 9A is the same type as in Fig. 1 in ref. (1) and is concave in the full range of H, indicating that the transition takes place in the all-or-none manner from state N to state D. The curve of Fig. 9C is the same type as in Fig. 3 in ref. (1) and is convex in the full range of H, showing that the transition is of the graded type as illustrated in Fig. 4 in ref. (1). These two curves clearly and quantitatively indicate the effect of the specificity of inter-unit

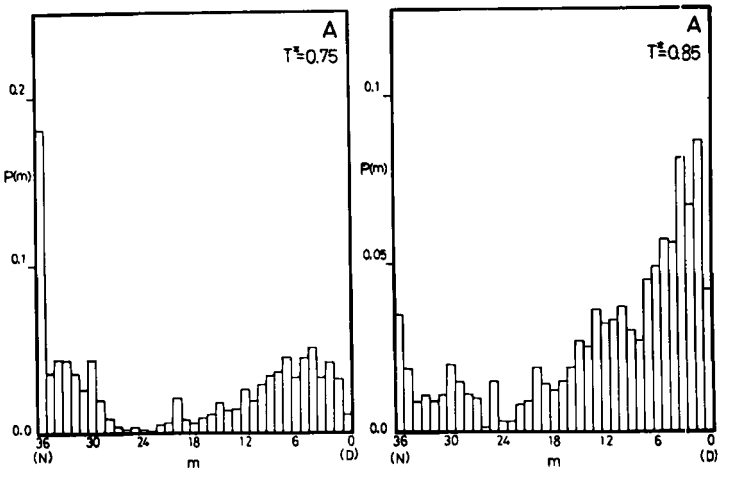

FIGURE 8
Population profiles near the transition temperature in case A, indicating that the transition is of the all-or-none type.

interactions on the statistical mechanical character of the transition. The case C of no specificity of inter-unit interactions can be regarded as the two-dimensional version of the three-dimensional lattice polymer with non-specific inter-unit interactions for the nearest neighbor units studied by Kron et al. (3). They also observed a globule-random coil transition of the graded type. Figs. 9A and C clearly demonstrate that an essential factor rendering the transition to be of the all-or-none type is the specificity of the inter-unit interactions. The curve of Fig. 9B for the intermediate specificity is the same type as the dotted line in Fig. 7 in ref. (1). A thermal conformational transition of the all-or-none type takes place at first from state N to state about $H = -13\varepsilon$. Then as the temperature is raised further, the peak position of the denatured state shifts gradually from $H = -13\varepsilon$ to $H = 0$. This case B probably simulates real globular proteins best out of the three cases of specificity studied in this paper.

The character of the transition is also reflected in the specific heat of the system near melting point. The specific heat, determined by energy fluctuation of the system, is given by

$$\frac{C}{k} = <(m - <m>)^2> / T^{*2} \qquad (5)$$

where $< \cdots >$ is the long-time average in the equilibrium state. The computed data are shown in Figs. 10A, B and C for the three cases of specificities A, B and C, respectively. Sharp peaks of the specific heat at melting point are observed in cases A and B, where the conformational transition from native to denatured states is of the all-or-none type. In case C, where the transition is of the graded type, there is no sharp peak. The shoulder appearing at the high temperature side of the peak in case B corresponds to the process of the gradual shift of the peak position in the denatured state. In this way the specific heat reflects a rather detailed process of the transition.

The time course of conformational changes starting from a denatured conformation is shown in Fig. 11 in the cases of specificities A and B for various temperatures. Most striking is the fact that in case A the specific native conformation is reached at low temperatures where the native conformation is stable. This result is important in two respects. One is that the Monte Carlo method employed in this paper is most effective for finding the native conformation of proteins. The other is that the system is assured of reaching the equilibrium state. These two aspects are discussed in turn in the following.

Predicting specific conformations of proteins from their amino acid sequences is undoubtedly one of the ultimate goals of protein research. So far, the method of minimization of conformational energy has been investigated for this purpose (4). It has become clear that this method involves the essential difficulty of multiple minima. A powerful algorithm must be found to overcome this difficulty and to find the global minimum on the multi-dimensional conformational energy surface. The Monte Carlo method of Metropolis et al.

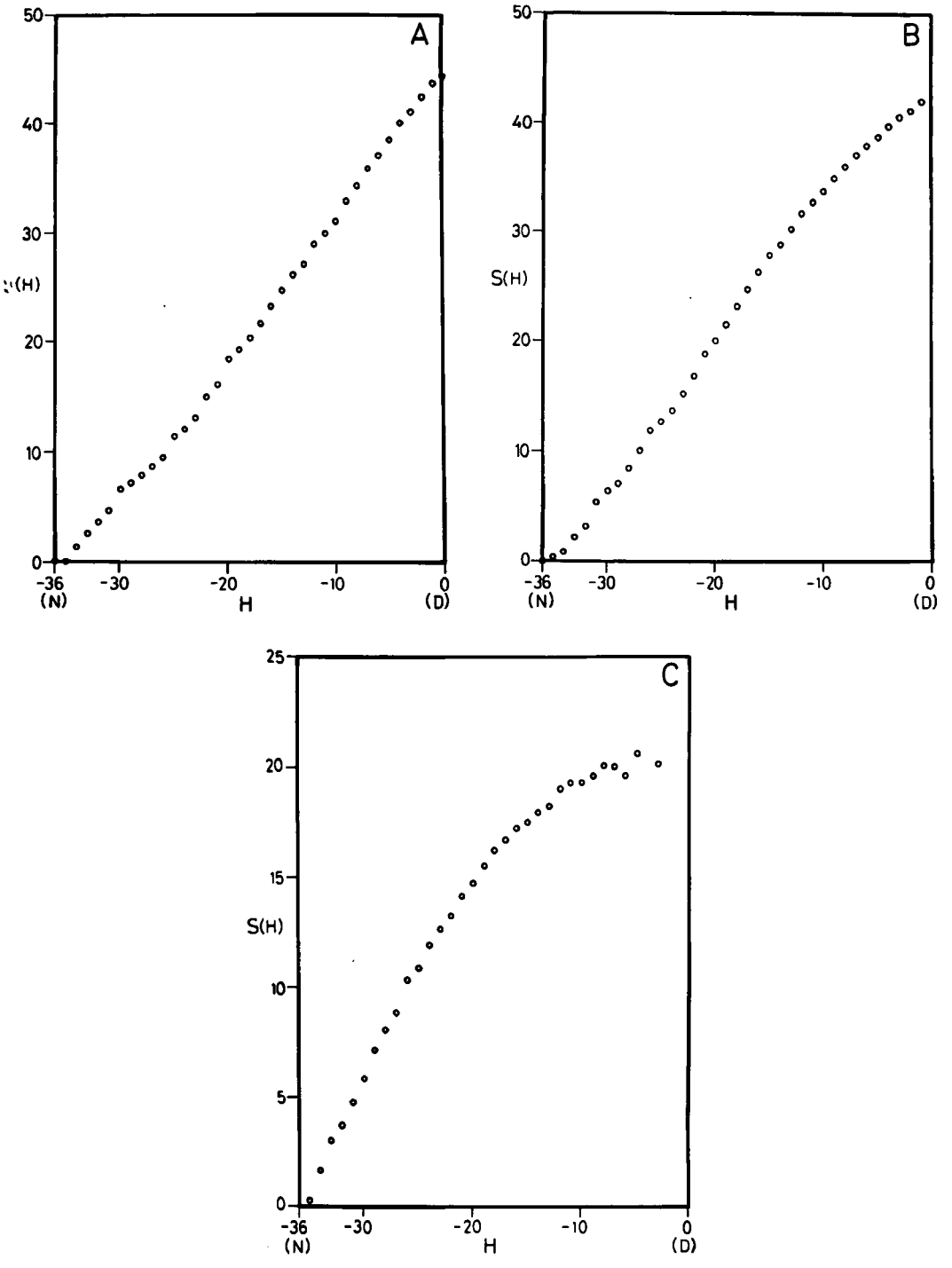

FIGURE 9

S-H curves for the three cases of specificity A, B and C.

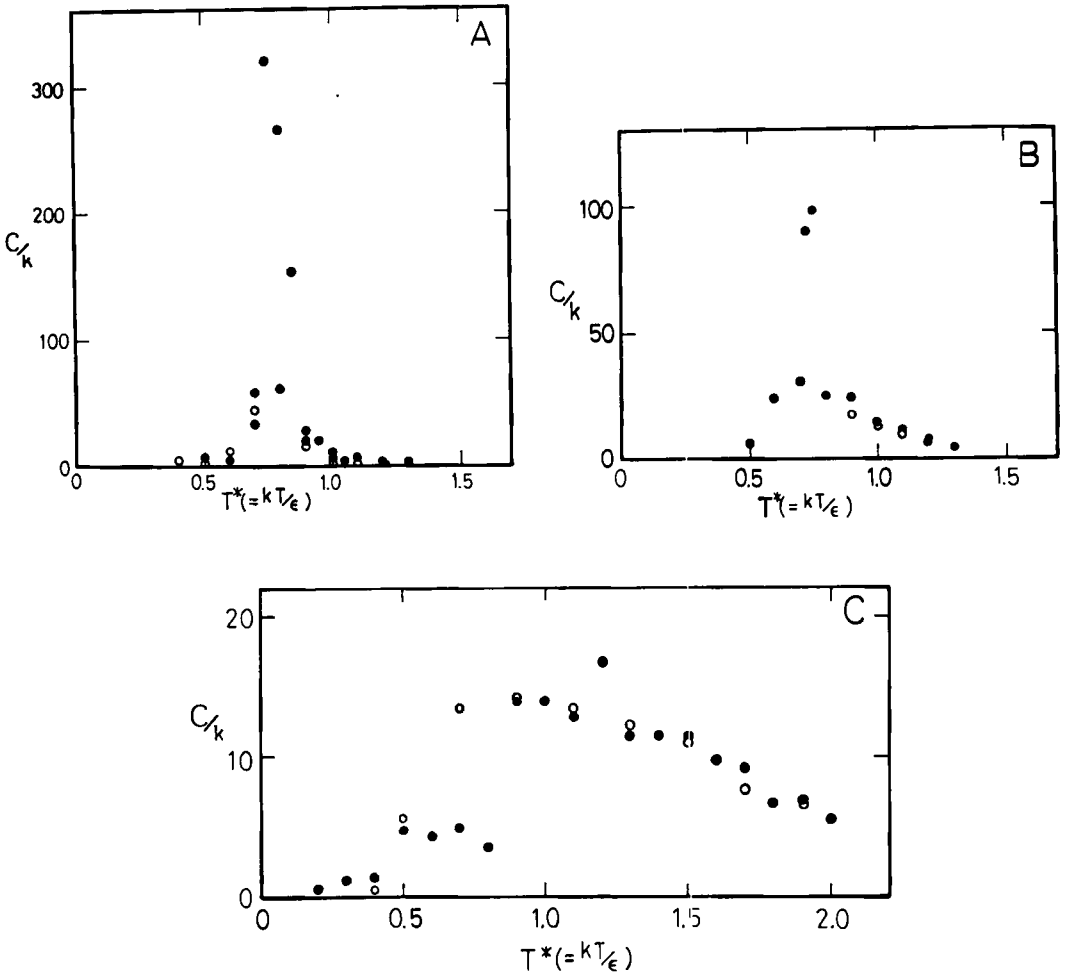

FIGURE 10

Specific heat of the system versus temperature T* for the three cases of specificity A, B and C.

appears to be a very hopeful one, since it has now been successfully applied to find the native conformation for the lattice model of proteins, no matter how simplified the model may be. As described in section II this method simulates the kinetic process fairly faithfully, implying that the application of the method may be understood to simulate the actual kinetic process of renaturation on the computer. It is interesting that the method simulating the real kinetic process is actually found to be effective. Fig. 12 shows several representative conformations in the course of renaturation in the case of A for $T^* = 0.4$. Many details about the processes of renaturation have not been discussed in this paper. However, it should be stressed that the study of conformations in the process of renaturation as shown in Fig. 12 can shed light on the role of nucleations etc.

The successful regeneration of the native conformation in the strong limit case of specificity A is interesting also from the point of view of assessing the role of the long-range interactions in the process of folding. Because of the recent recognition of the correlations between short-range amino acid sequence and secondary struc-

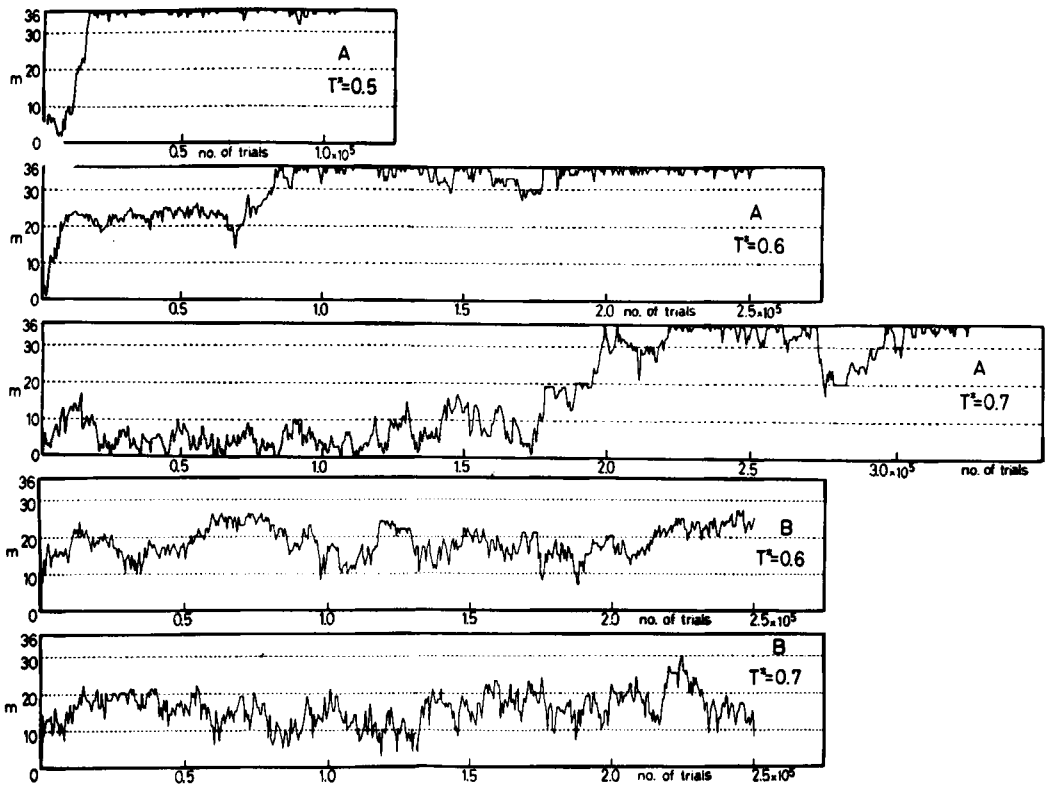

FIGURE 11

Courses of conformational changes starting from a denatured conformation in cases A and B for various temperatures.

tures, it is currently believed that the short-range interactions play an important role in the folding process. However, the present observation indicates that the role of the short-range interactions is to be assessed only in relation to the role of the long-range interactions, because regeneration of the native conformation was possible in the model only with long-range interactions.

On the other hand, Fig. 11 also indicates that the Monte Carlo method employed in this paper is not successful for finding the native conformation in the more realistic case of intermediate specificity B. This observation indicates that the effectiveness of the Monte Carlo method is not independent of the intramolecular interactions. At present we do not consider the impossibility of regenerating the native conformation in the case of specificity B as an indication of the inadequacy of the Monte Carlo method; instead we regard it as an indication of the inadequacy of the intramolecular interactions in the model. We are trying to incorporate a new feature into the model whereby the renaturation to the native conformation is facilitated. One possibility is the incorporation of the short-range interaction discussed in section I.

The basic theory of the Monte Carlo method of Metropolis et al. assures that the long-time average converges to the equilibrium value. Time is said to be long or short in comparison with the relaxation times of the system. Relaxation times are, in general, not known beforehand. This creates an inherent ambiguity in the simulation method of phenomena involving a phase transition or a phase-transition-like change, where relaxation times become very long. It is difficult to exclude all ambiguities from the interpretation of data of simulations, especially at or near the

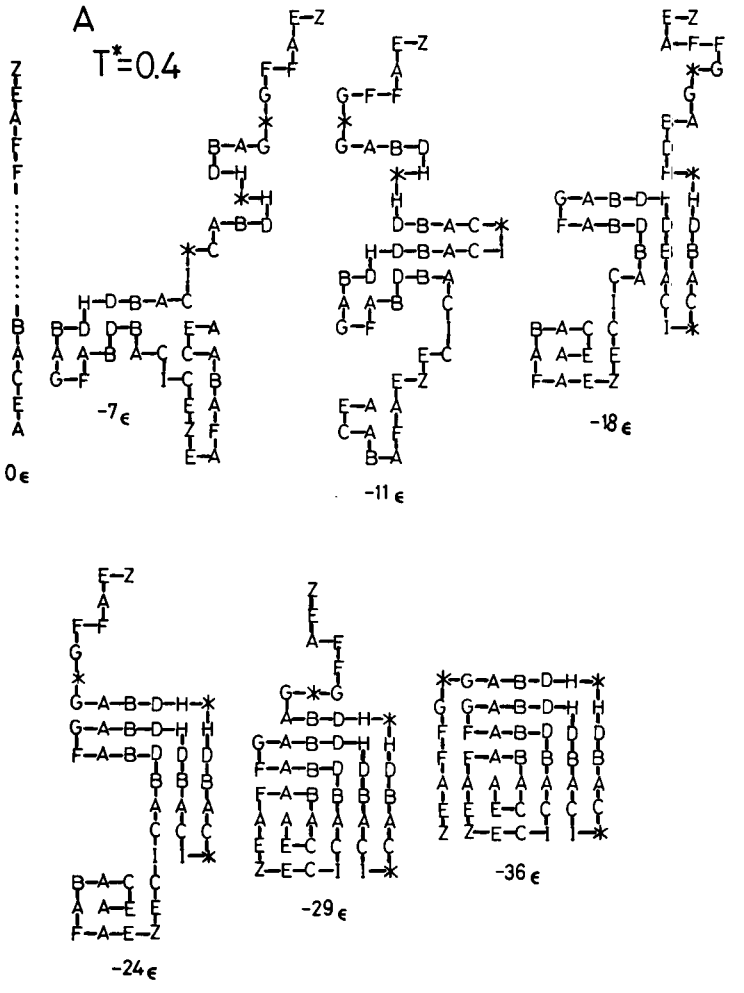

FIGURE 12
Several representative conformations in the course of renaturation in the case of A for T* = 0.4.

transition point. In the case of specificity A, where renaturation to the native conformation does actually take place at low temperatures, the equilibrium state appears to be reached within the trial numbers shown in Figs. 6 and 11. In other words, the relaxation time is not very long in the strong limit case of specificity A. In the weak limit case of specificity C, where the transition is of the graded type, the relaxation time is not very long, either. In the intermediate specificity B the relaxation time appears to be very long. In this case the protein remains fluctuating about the native conformation up to T* = 0.7, when started from the native conformation. A sudden denaturation takes place as the temperature is raised to T* = 0.725. These observations indicate that the native state is stable up to about T* = 0.7. When started from an extended conformation (chosen from conformations in the denatured state), the renaturation to the native conformation is not observed at temperatures below T* = 0.7 within the trial numbers tried in Fig. 11. The chain polymer is trapped in a state which apparently corresponds to a local minimum on the conformational energy surface. The trapped states, which do not represent the equilibrium state, are not included in Figs. 7B and 10B. Further study is now under way, based on the view that these trappings do not occur in real protein due to factors not yet incorporated in the model of specificity B.

The study in this series of papers has three

features; (a) a model of proteins, (b) a method which allows the model to move in simulation of the kinetic processes of proteins, and (c) to reveal the mechanism of the kinetic processes by applying (b) to (a). In these terms the purpose of this series of papers can be re-stated as follows. To construct a model with essential factors in as simple a form as possible, and to develop a method by which the kinetic behavior of the model can be studied so that the essence of the statistical mechanical properties of conformations of proteins, especially folding, unfolding and fluctuations, can be understood. There is a deficiency between the present lattice model and real proteins, especially in the way the heterogeneity of the amino acid sequence is expressed by the specificity of inter-unit interactions. Thus, the findings in this paper can not yet be utilized directly to predict the native conformation of proteins from the amino acid sequence. However, in this series of papers we hope to reveal which aspects of real proteins are essential for the process of folding and thus can be used in the algorithm for predicting the native conformation.

IV. CONCLUSIONS

A lattice model of proteins with specific inter-unit interactions was shown to provide a good and effective method for studying the statistical mechanical properties of conformations of globular proteins. Three cases of the specificities A, B and C, specifying the inter-unit interactions of strong limit, intermediate, and weak limit, respectively, were studied. In the cases of the strong limit and intermediate specificities, denaturations were found to be of the all-or-none type, while in the case of the weak limit specificity, denaturation was found to be of the graded type. It thus follows that the specificity of inter-unit interactions reflecting the inhomogeneity in the amino acid sequence is essential for the all-or-none character of the denaturation of globular proteins. The lattice model is also expected to be effective for elucidating fluctuations in native and denatured state and the detailed processes of denaturation and renaturation. The lattice model was studied by the Monte Carlo method of Metropolis et al. This method was able to approximately simulate a kinetic process. In the strong limit case of specificity A renaturation to the native conformation was observed actually to take place, indicating the effectiveness of the Monte Carlo method for predicting the native conformations of globular proteins. In the intermediate case of specificity B, a more realistic case than A, renaturation to the native conformation was hindered, because the conformational state was trapped in a local minimum on the conformational energy surface. Incorporation of new features into case B such as emphasized short-range interactions, and so forth, is currently being tried to see if such a change in the model can help realize the renaturation to the native conformation.

ACKNOWLEDGMENTS

Critical reading of the manuscript by Dr. N. Sakamoto is gratefully acknowledged. This work has been partially supported by grants-in-aid from the Ministry of Education, Japan [Sogokenkyu A, No. 838049 (1973), No. 938045 (1974); Ippankenkyu C, No. 958073 (1974)] and by a cooperative research program of the Research Institute for Fundamental Physics, Kyoto University. The computation was performed at the Computer Center, Kyushu University.

REFERENCES

1. Gō, N. (1975) *Int. J. Pept. Prot. Res.* **7**, 313–323
2. Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. (1953) *J. Chem. Phys.* **21**, 1087–1092.
3. Kron, A. K., Ptitsyn, O. B., Skvortsov, A. M. & Fedorov, A. K. (1967) *Molekulyarnaya Biologiya* **1**, 576–582.
4. Scheraga, H. A. (1968) *Adv. Phys. Org. Chem.* **6**, 103–184.

Address:

Nobuhiro Gō
Department of Physics
Faculty of Science
Kyushu University
Fukuoka 812
Japan