Biophysical Chemistry 6 (1977) 117–130


# STRUCTURAL CHANGES AND FLUCTUATIONS OF PROTEINS.
## I. A statistical thermodynamic model

Akira IKEGAMI
*Department of Physics, Faculty of Science, University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113, Japan*

Received 3 August 1976

A general theory of the structural changes and fluctuations of proteins has been proposed based on statistical thermodynamic considerations at the chain level.

The "structure" of protein was assumed to be characterized by the state of secondary bonds between unique pairs of specific sites on peptide chains. Every secondary bond changes between the bonded and unbonded states by thermal agitation and the "structure" is continuously fluctuating. The free energy of the "structural state" that is defined by the fraction of secondary bonds in the bonded state has been expressed by the bond energy, the cooperative interaction between bonds, the mixing entropy of bonds, and the entropy of polypeptide chains. The most probable "structural state" can be simply determined by graphical analysis and the effect of temperature or solvent composition on it is discussed. The temperature dependence of the free energy, the probability distribution of structural states and the specific heat have been calculated for two examples of structural change.

The theory predicts two different types of structural changes from the ordered to disordered state, a "structural transition" and a "gradual structural change" with rising temperature. In the "structural transition", the probability distribution has two maxima in the temperature range of transition. In the "gradual structural change", the probability distribution has only one maximum during the change.

A considerable fraction of secondary bonds is in the unbonded state and is always fluctuating even in the ordered state at room temperature. Such structural fluctuations in a single protein molecule have been discussed quantitatively.

The theory is extended to include small molecules which bind to the protein molecule and affect the structural state. The changes of structural state caused by specific and non-specific binding and allosteric effects are explained in a unified manner.

## 1. Introduction

The most significant characteristic of proteins is their highly ordered structure in solution which far exceeds that usually observed with synthetic polypeptides.

The tertiary and quaternary structure of proteins is maintained by many kinds of secondary bonds between intra- or inter-peptide chains. The bond energy due to secondary bonds such as hydrogen bonds, hydrophobic interactions, electrostatic or dipole interactions is perhaps a few times that of thermal energy, however, and if there are no cooperative actions between the secondary bonds they may be broken frequently by the thermal agitation. Therefore, cooperative interaction appears essential to maintain the tertiary structure of proteins.

As an upper limit in describing cooperativity, a simple two state model is usually presented to explain a number of structural changes of proteins [1–5]. The allosteric effects of enzymes have also been explained by two state models [6,7]. Because of the simplicity of the model, however, its analysis provides little information on the molecular characterization of proteins.

Although the cooperative interaction is enhanced by the polypeptide chain on which the bonding sites of secondary bonds are specifically arranged, it is unlikely that the cooperative effects are sufficient to cover all the secondary bonds in a protein as predicted by the simple two state model and local fluctuations in structure are expected even in a highly ordered native state.

In fact, the presence of local fluctuations was experimentally confirmed using the hydrogen exchange methods [8,9], and it is now necessary to consider more precise and realistic theories based on the statistical considerations at the chain level. Recently Hermans et al. [10] proposed a multi-state model and treated the folding and unfolding of proteins. In their model, however, the relation between the model and molecular properties such as the entropy of the chain are obscure.

The present series of papers proposes a general theory for the structural changes and fluctuations in proteins and applies the theory to real systems. In the first paper the basic concepts and general features of the theory are presented. The model can reasonably explain a number of phenomena related to structural changes and fluctuations in proteins. For this reason the model can aid in our understanding of the nature of proteins at the molecular level. In the following paper, the denaturation of several globular proteins are analyzed quantitatively with special attention to the effect of water structure.

## 2. Description of the model

(1) In an "ideal" solution at 0K, protein molecules take a unique conformation dictated by the law of minimum energy. A number of intra- or inter-chain secondary bonds between unique pairs of specific sites contribute to form the fixed conformation. Therefore, there is no flexibility of chains and the entropy of the system equals zero.

(2) Each secondary bond can exist in either a bonded or unbonded state. At 0K, all secondary bonds are in the bonded state and as the temperature rises, the fraction of these in the bonded state Y in a molecule decreases and the flexibility of chains increases.

(3) A bonded state can be formed only between a unique pair of specific sites. The number of possible bonds $N_0$ is independent of temperature because many different kinds of binding sites are uniquely arranged on a chain and they can not redistribute independently to make other bonding forms possible. Although it may be possible that a bond is formed between other pairs of sites, we neglect such cases. The probability of forming alternative bonds is likely to be quite small since neighboring sites would probably be in less favourable conditions than in the case of the unique pair.

(4) There is a loss of energy when neighbouring bonds are in different states. That is, cooperative interaction between neighbouring bonds exists since the binding sites are connected to each other by the polypeptide backbone.

(5) As a result of thermal agitation, every secondary bond fluctuates in a random way between the bonded and unbonded state, but with constant average rates. Thus, the fraction of secondary bonds in their bonded state $Y$ can not be constant but is continuously fluctuating. The average fraction of bonded state $\overline{Y}$, however, can be determined by minimizing the free energy of the system. For each molecule, the average fluctuation of this fraction, $\Delta Y^2$, can also be determined by the free energy change near its minimum point.

(6) At a certain "structural state" which is characterized by a specific value of $Y$ or $X$ ($\equiv 2Y-1$), a protein molecule can assume different "structures" and can fluctuate between these "structures". The term "structure" describes hereafter a group of conformations in which the state of all the secondary bonds are specified. Therefore, many different conformations belong to a structure because of the flexibility of polypeptide chains. Moreover, it is worth noting that most measurable quantities relate to structures rather than conformations.

## 3. Free energy

In order to simplify the calculations we make the following assumptions.

(1) All of the possible secondary bonds, the number of which is $N_0$, have the same properties, that is the same bond energy $\epsilon$ and the same number of nearest neighbouring bonds $Z$; the same energy loss $J$ arises for any two nearest neighbour bonds which are in a different state.

Then the state of a bond in a protein molecule can be assumed to behave as an Ising spin state in a lattice topologically similar to the protein.

(2) The flexibility of the main and the side chains increases as secondary bonds are broken. The number of conformations produced when a single bond breaks is the same for all bonds. Then, the chain entropy of a molecule is proportional to the number of unbonded

state $N_u$. If the standard state of energy is set at 0K, the Gibbs free energy of a protein molecule is given by

$$\begin{aligned} G(T,X) &= H - ST \\ &= N_0\epsilon\tfrac{1}{2}(1-X) + N_0 ZJ\tfrac{1}{4}(1-X^2) \\ &+ N_0 kT[\tfrac{1}{2}(1+X)\ln\tfrac{1}{2}(1+X) + \tfrac{1}{2}(1-X)\ln\tfrac{1}{2}(1-X)] \\ &- N_0\tfrac{1}{2}(1-X)\alpha T , \end{aligned} \tag{1}$$

where $k$ is the Boltzmann constant and parameter $X$ which characterizes "structural state" is related to the fraction of secondary bonds in bonded state $Y$ by

$$X = 2Y - 1$$

or

$$\begin{aligned} \tfrac{1}{2}(1+X) &= Y = N_b/N_0 , \\ \tfrac{1}{2}(1-X) &= 1 - Y = N_u/N_0 . \end{aligned} \tag{2}$$

The first term of eq. (1) expresses the total bond energy. The second term represents the cooperative contribution of forming secondary bonds between nearest neighbour bonds and the third term the mixing entropy of bonds. The fourth term is the entropy of the chain in which $\alpha$ is the chain entropy per secondary bond produced by the breaking of all bonds. For simplicity we put

$$\epsilon/ZJ = A, \quad 2k/ZJ = B, \quad \alpha/2k = C . \tag{3}$$

Eq. (1) reduces to

$$\begin{aligned} G(T,X) &= N_0\tfrac{1}{2}ZJ[A(1-X) + \tfrac{1}{2}(1-X^2) \\ &+ BT[\tfrac{1}{2}(1+X)\ln\tfrac{1}{2}(1+X) + \tfrac{1}{2}(1-X)\ln\tfrac{1}{2}(1-X)] \\ &- (1-X)BTC] . \end{aligned} \tag{4}$$

## 4. Minimum free energy and thermal effects on the structural state

The equilibrium distribution of the "structural state" of the protein is determined by the free energy. A most probable value approximation is made in which the average value of $X$, $\bar{X}$, is set equal to $X_m$, the most probable value of $X$.

By the condition of minimum free energy

$$\frac{\partial G(T,X)}{\partial X} = 0 , \tag{5}$$

the most probable value $X_m$ is given by the following equation,

$$N_0\tfrac{1}{2}ZJ[-A - X_m + BT\tfrac{1}{2}\ln\left(\frac{1+X_m}{1-X_m}\right) + BTC] = 0. \tag{6}$$

Eq. (6) can be solved graphically after rearranging to

$$\tanh^{-1} X_m \equiv \tfrac{1}{2}\ln\left(\frac{1+X_m}{1-X_m}\right) = \frac{A+X_m}{BT} - C . \tag{7}$$

Then, by letting

$$\frac{A+X}{BT} - C = x , \tag{8}$$

the value of $X_m$, or the most probable structural state, is determined from the intersection of the straight line expressed by eq. (8) and the hyperbolic tangent curve expressed by $\tanh x = X$ (see fig. 1).

As shown in fig. 1, the straight lines expressed by eq. (8) pass through the point $p(-C, -A)$ and have a slope $BT$ varying with temperature. Since the parameters $A$, $B$ and $C$ depend only on the protein species considered, point p will be called the "characteristic point" of a protein.

The number of intersections between the straight line and the hyperbolic tangent curve depends on the temperature and the position of the characteristic

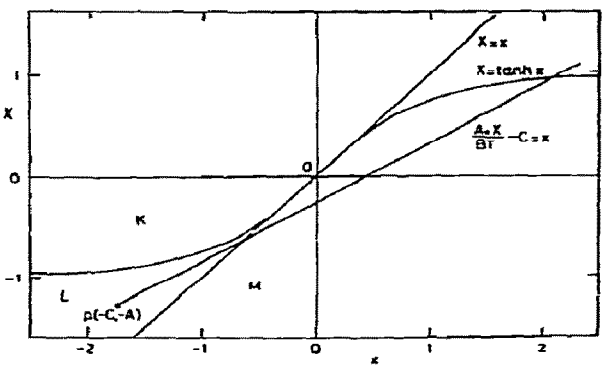


Fig. 1. The graphic determination of the most probable structural state $X_m$. The most probable value $X_m$ can be determined from the intersection of the hyperbolic tangent curve $X = \tanh x$ and the straight lines expressed by eq. (8) which pass through the characteristic point $p(-C, -A)$ and have slope $BT$.

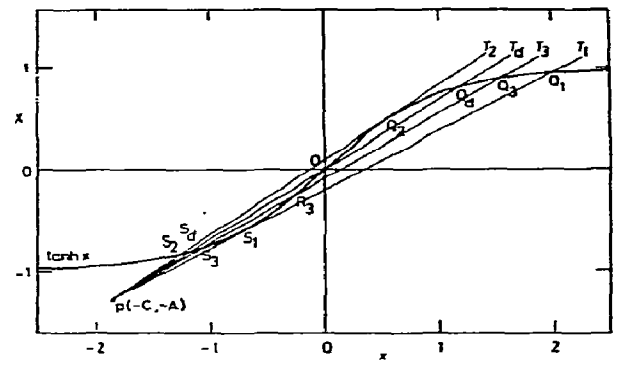


Fig. 2. The graphic analysis of the change in the most probable structural state $X_m$ with temperature (see text).

point. For example, as shown in fig. 2, if the characteristic point p is in the regions K and L in fig. 1, (where $C > A$), only one intersection occurs at a temperature lower than $T_1$ or at temperature higher than $T_2$ and two intersections at $T_1$ and $T_2$; there are three intersections between these temperatures.

At 0K, the slope of the straight lines takes the value +0 and they intersect at infinity ($x = \infty$, $X = 1$) where the completely ordered state is found. As the temperature rises, the value of $X_m$ decreases to a small extent but the most probable structural state of the protein remain almost in the "ordered state". At a certain temperature, $T_3$, above $T_1$, the straight line intersects with the curve at three points $Q_3$, $R_3$ and $S_3$ and the free energy of the system has two minima at $Q_3$ and $S_3$ and a maximum at $R_3$. When the temperature $T_3$ is lower than the denaturation temperature $T_d \equiv A/BC$, the free energy corresponding to $Q_3$ is lower than that corresponding to $S_3$ and the "ordered state" is more favourable than the "disordered state". At $T_d$, where the straight lines pass through the point 0, the free energy at both minima has the same value and above $T_d$ the free energy corresponding to $S_3$ becomes lower than that corresponding to $Q_3$. Thus, at the denaturation temperature $T_d$, the most probable structural state changes discretely from the ordered state to the disordered state and a "structural transition" occurs, which is essentially a first order transition.

Especially when the characteristic point is in the region K in fig. 1, where $C > \tanh^{-1} A$, after a sharp transition from the ordered to the disordered state, to some extent secondary bonds reform in the disordered state with a further rise in temperature. In some proteins [2,11] evidence for this effect is given by a small change in the ultraviolet absorption spectrum which is usually attributed to a solvent effect.

In contrast, if the characteristic point is in the region M of fig. 1, where $A \geqq C$, the straight line and the hyperbolic tangent curve always intersect at only one point and the value of $X_m$ decreases gradually with rising temperature. The phenomenon will be referred as a "gradual structural change". In the special case of $A = C$, the "gradual structural change" can be assumed to be a second order transition.

Variations in $X_m$ with temperature are shown in figs. 3 and 4. These figures illustrate the fact that the protein may exist in two more-or-less separate states and a sharp transition takes place at a certain temperature if the value of $C/A$ is larger than unity.

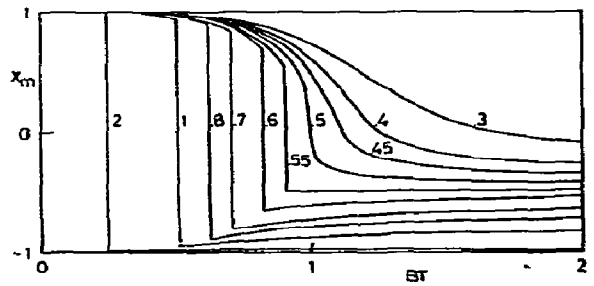


Fig. 3. The most probable structural state $X_m$ as function of $BT$ at various values of $C$, with $A = 0.5$.

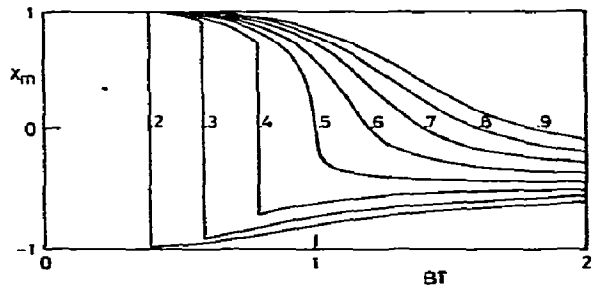


Fig. 4. The most probable structural state $X_m$ as function of $BT$ at various values of $A$, with $C = 0.5$.

## 5. The free energy near the transition point and the probability distribution

In the previous section we discussed the effect of temperature on the structural state only through changes in $X_m$, the fraction of secondary bonds in the bonded state at the minimum free energy point. When the number of bonds is very large, as in the case of a crystal, the value $X_m$ correctly expresses the average value $\bar{X}$ of the real system and fluctuation from the mean value is negligible.

In the case of proteins, however, the total number of bonds $N_0$ is expected to be of the order only of a hundred or so, and the most probable value $X_m$ does not necessarily represent the average value $\bar{X}$ for a protein molecule near the transition point. Hence, the conformational fluctuations are expected to be very large. The transition of the average structural state of a protein molecule, which is essentially a first order transition, is expected to be broad.

The free energy of a typical case undergoing a "structural transition" has been calculated and is plotted against $X$ in fig. 5. As shown in the figure, the free energy has two minima corresponding to the "ordered" and "disordered" state of the previous discussion. At low temperatures, the free energy at the minimum corresponding to the "ordered state" is lower than that corresponding to the "disordered state". With rising temperature, the difference between the free energies of the two minima decreases and becomes zero at the transition temperature. At temperatures above the transition temperature, the free energy of the "disordered state" is lower than that of the "ordered state".

The probability density $\rho(T, X)$ of the system has been calculated by

$$\rho(T, X) = \frac{e^{-G(T, X)/kT}}{\int_{-1}^{1} e^{-G(T, X)/kT}\, dX} \tag{9}$$

and is shown in fig. 6. Then, the average value $\bar{S}(T)$ of a certain physical quantity $S(T, X)$ can be calculated by

$$\bar{S}(T) = \int_{-1}^{1} S(T, X)\, \rho(T, X)\, dX\,. \tag{10}$$

The average structural state $\bar{X}(T)$ has been calculated and is shown in fig. 7.

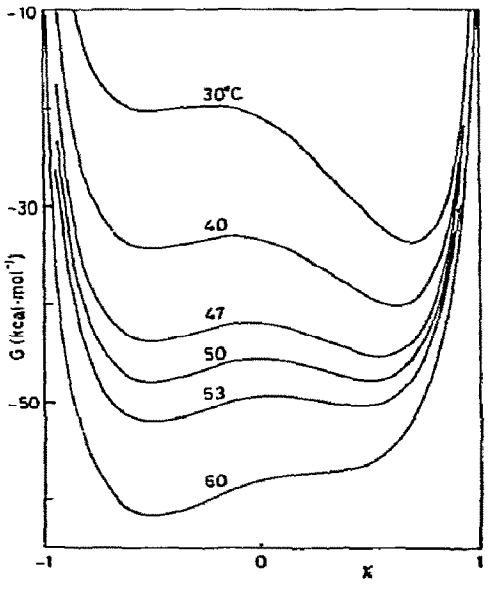


Fig. 5. The Gibbs free energy for a typical case undergoing a "structural transition" at various temperatures calculated according to eq. (4), with $A = 0.5$, $B = 2.813 \times 10^{-3}$, $C = 0.55$, $N_0 = 500$. These values correspond to $T_d = 50°C$, $\epsilon = 706$ cal/mole, $ZJ = 1412$ cal/mole and $\alpha = 2.185$ eu.

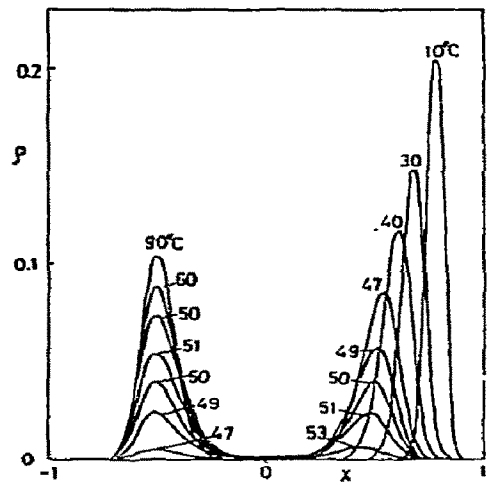


Fig. 6. The probability density $\rho(T, X)$ at various temperatures calculated according to eq. (9) with the same values of the parameters as in fig. 5.

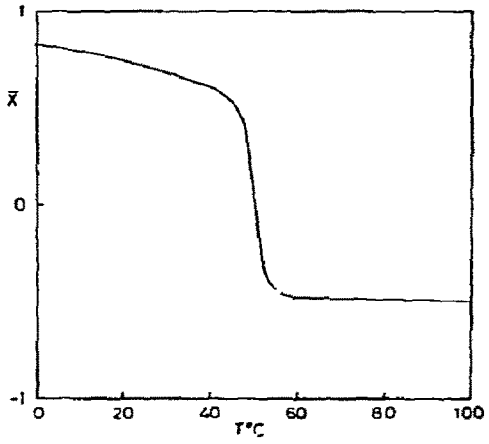


Fig. 7. The change in the average structural state $\bar{X}$ with temperature calculated according to eq. (10) with the same values of the parameters as in fig. 5.

Near the transition temperature, the difference in free energy between the two states is not very large; hence the probability distribution curve has two completely separate peaks as shown in fig. 6. In this temperature range, each protein molecule takes either the "ordered" or "disordered" state separately. The average value $\bar{X}$, which is apparently different from the most probable value $X_m$, does not change so sharply in this temperature range.

The maximum free energy difference separating two states (at $X = 0$) is probably only a few or ten kcal as illustrated in fig. 5. The rate of the structural transition between two states may be expected to be fast. Such fast transition rates are observed in many proteins. That is, the free energy maximum between ordered and disordered states is large enough to separate the "two states" but low enough to permit a fast transition rate, this being the physical condition for equilibrium between two different states assuming no specific effect from the solvent. The origin of this condition comes from the fact that the number of bonds or amino acids in protein molecules has an appropriate value.

In fig. 8, the free energy of a system undergoing a "gradual structural change" is plotted against $X$. The probability distribution $\rho(X)$ and average value $\bar{X}$ are also shown in figs. 9 and 10. In this case, contrary to the case of the "structural transition", the probability distribution has only one peak for all temperatures,

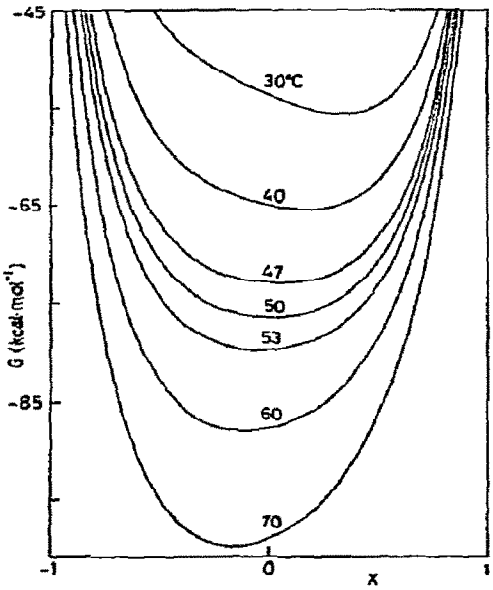


Fig. 8. The Gibbs free energy for a typical case undergoing a "gradual structural change" at various temperatures calculated according to eq. (4), with $A = 0.5$, $B = 3.438 \times 10^{-3}$, $C = 0.45$, $N_0 = 500$. These values correspond to $T_d = 50°C$, $\epsilon = 578$ cal/mole, $ZJ = 1156$ cal/mole and $\alpha = 1.609$ eu.

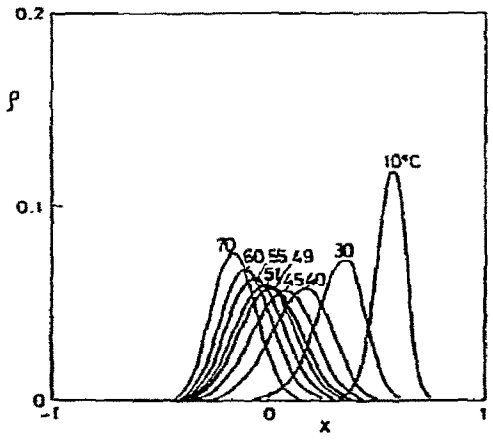


Fig. 9. The probability density $\rho(T, X)$ at various temperatures calculated according to eq. (9) with the same values of the parameters as in fig. 8.

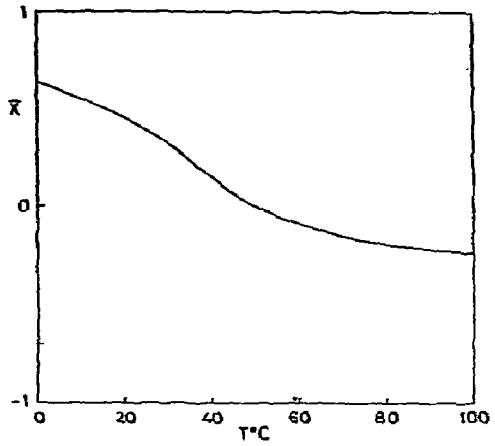


Fig. 10. The change in the average structural state $\bar{X}$ with temperature calculated according to eq. (10) with the same values of the parameters as in fig. 8.

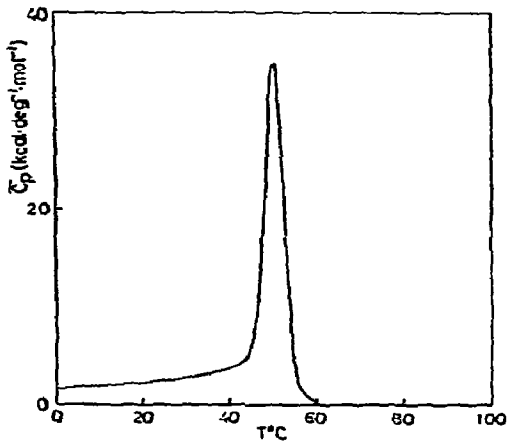


Fig. 11. The change in the average heat capacity $\bar{C}_p$ with temperature calculated according to eq. (13) with the same values of the parameters as in fig. 5.

and the protein cannot be described by "two states" even though the average value $\bar{X}$ changes markedly in a certain range of temperature.

## 6. Heat capacity

Since changes in enthalpy or heat capacity are directly related to changes in the fraction of bonded states, their measurement is important to characterize changes in structural state. As the enthalpy of the system is given by

$$H = -T^2 \frac{\partial(G/T)}{\partial T} \approx N_0 \epsilon \tfrac{1}{2}(1-X) + N_0 Z J \tfrac{1}{4}(1-X^2), \quad (11)$$

the average enthalpy $\bar{H}(T)$ and heat capacity $\bar{C}_p$ are given by

$$\bar{H}(T) = N_0 \epsilon \tfrac{1}{2}(1-\bar{X}) + N_0 Z J \tfrac{1}{4}(1-\overline{X^2}) \quad (12)$$

and

$$\bar{C}_p(T) = \frac{d\bar{H}}{dt} = -\tfrac{1}{2} N_0 \epsilon \frac{d\bar{X}}{dT} - \tfrac{1}{4} N_0 Z J \frac{d\overline{X^2}}{dT}. \quad (13)$$

The changes in average heat capacity $\bar{C}_p$ with temperature for the two examples in the previous sections are shown in figs. 11 and 12. The shape of the curves representing the change in heat capacities of proteins measured near their transition [12] is intermediate

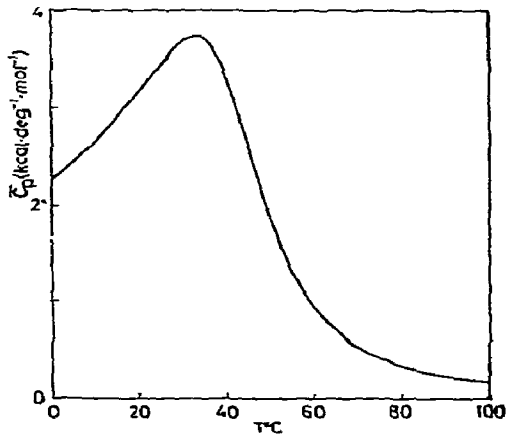


Fig. 12. The change in the average heat capacity $\bar{C}_p$ with temperature calculated according to eq. (13) with the same values of the parameters as in fig. 8.

between that of these two figures.

## 7. Measurable quantities, local structure and the "apparent" intermediate state

Most measurable quantities in a protein, such as its ultra violet absorption spectrum, optical rotatoty

power, or binding constant of a substrate are expected to depend on the corresponding "local structure" of the protein. Here we define the term "local structure" as a group of conformations in which the state of a certain set ($Q_s$) of $N_s$ secondary bonds is specified. The state of other secondary bonds is not specified. A "local structural state" is characterized by the fraction of bonded states $Y_s$ in the set $Q_s$.

Although the value of a measurable quantity $s$ could differ from one local structure to another in a particular set $Q_s$, we assume simply that $s$ has either one of the two values $s_1$ or $s_2$ in all local structures in the set $Q_s$; the value $s_1$ corresponds to the local structural states characterized by $Y_s \geqq Y_s^0$ and $s_2$ corresponds to that characterized by $Y_s < Y_s^0$.

In the approximation used in the present theory, the probability $P_{Y_s}^{N_s}$ of a local structural state characterized by $Y_s$ is given simply by

$$P_{Y_s}^{N_s} = \frac{(N_0 - N_s)C_{(N_0 Y - N_s Y_s)} \times_{N_s} C_{N_s Y_s}}{{}_{N_0}C_{N_0 Y}}. \tag{14}$$

Similarly the average value of $s$, $\bar{s}(Y)$ at a structural state characterized by $Y$ is

$$\bar{s}(Y) = \frac{s_1 \int_{Y_s^0}^{1} P_{Y_s}^{N_s} \, dY_s + s_2 \int_{0}^{Y_s^0} P_{Y_s}^{N_s} dY_s}{\int_{0}^{1} P_{Y_s}^{N_s} \, dY_s}. \tag{15}$$

The average value of $\bar{s}(Y)$ over all structural states $\langle \bar{s} \rangle$ which is the only value that can be obtained experimentally is given by

$$\langle \bar{s} \rangle = \int_{-1}^{1} \bar{s}(Y) \rho(T, X) \, dX . \tag{16}$$

For the quantity $s$, we can substitute various measurable quantities such as optical rotatory power, intensity of ultraviolet absorption, the binding constant of substrate or inhibitors and the activity of the enzyme.

For example, the ultraviolet absorption of the chromophore A represented schematically in fig. 13 must be related to the local structure near it. When bond a is either in a bonded or unbonded state, chromophore A is expected to have either an absorption intensity $s_1^A$ or $s_2^A$ respectively. In this case, $N_s = 1$, $P_1^1 = Y$, then $\overline{s^A}(Y)$ which is a linear function of $Y$, is given by

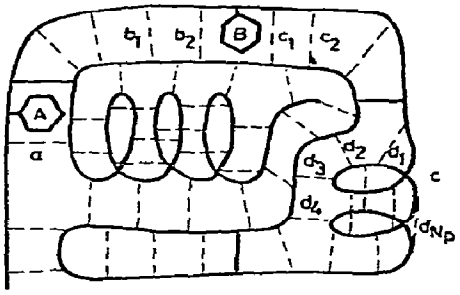


Fig. 13. A schematic representation of a protein structure.

$$\overline{s^A}(Y) = s_1^A Y + s_2^A (1-Y) . \tag{17}$$

On the other hand, consider chromophore B which has an absorption intensity $s_1^B$ when one or both of the two bonds $b_1$, $b_2$ and one or both of the two bonds $c_1$, $c_2$ are in the bonded state, and which otherwise has an absorption intensity $s_2^B$. The absorption intensity of B depends on the states of two sets of bonds $b_1$, $b_2$ and $c_1$, $c_2$.

The probability $P_{Y_s}^2$ for the set of two bonds is approximately given by $P_1^2 = Y^2$, $P_{0.5}^2 = 2Y(1-Y)$ and $P_0^2 = (1-Y)^2$ if the cooperativity between bonds can be neglected. Then $\overline{s^B}(Y)$ is given by

$$\overline{s^B}(Y) = s_1^B Y^2 (2-Y)^2 + s_2^B [(1-Y)^2 (1+2Y-Y^2)] . \tag{18}$$

The helical structure C in fig. 13 is expected to occur when there are more than $N_p Y_p^0$ bonds out of the possible $N_p$ bonds ($d_1, d_2, \ldots, d_{N_p}$) in the region near the helical structure. Then the average value of optical rotatory power for this helix is given by the same equation as (15).

More complicated quantities such as the binding constant or the enzyme activity must also be related to a certain number of local structural states as in the case of chromophore B.

It should be noted here, however, that the expression for $P_{Y_s}^{N_s}$ as given by eq. (14) becomes invalid as the size of local structure is decreased because of the cooperative interaction between bonds. If the cooperativity between bonds is sufficient to include all $N_s$ bonds, the function $\bar{s}(Y)$ must be directly proportional to $Y$. When the size of the local structure or the

value of $N_s$ is large, however, eq. (15) shows that $\bar{s}(Y)$ approaches a step function that increases from $s_2$ to $s_1$ at $Y_s = Y_s^0$.

Two chromophores of the same species that have different numbers of secondary bonds associated with them will also have different $s(Y)$ functions. It is possible that an "apparent intermediate" state of the protein occurs even when the stable structural state has no intermediate as is the case of structural transition. The same situation is expected for the optical rotatory power of proteins in which two or more different sizes of helices are present. Therefore, the finding of an "apparent" intermediate state does not necessarily indicate the existence of a true intermediate state.

## 8. Structural fluctuations in a single protein molecule

### *8.1. Effects of solvent fluctuations*

In the preceding sections, we assumed that a protein molecule is in thermal equilibrium with a solvent at constant temperature $T$, pressure $P$ and composition $N_i$. In a true solution, these thermodynamic quantities fluctuate since the volume of solvent with which a single protein molecule is in thermal equilibrium is probably small.

In general, the average fluctuation of the structural state of the protein $\overline{\Delta X^2} = \overline{(X - X_m)^2}$, due to fluctuation of $T, P, N_i$, in this solvent is expressed by

$$\overline{\Delta X^2} = \left(\frac{\partial X}{\partial T}\right)^2 \overline{\Delta T^2} + \left(\frac{\partial X}{\partial P}\right)^2 \overline{\Delta P^2} + \sum_{ij} \bar{Z} \left(\frac{\partial X}{\partial N_i}\right)\left(\frac{\partial X}{\partial N_j}\right) \overline{\Delta N_i \Delta N_j} + \left(\frac{\partial X}{\partial T}\right)\left(\frac{\partial X}{\partial P}\right) \overline{\Delta T \Delta P} + \sum_i \left(\frac{\partial X}{\partial T}\right)\left(\frac{\partial X}{\partial N_i}\right) \overline{\Delta T \Delta N_i} + \sum_i \left(\frac{\partial X}{\partial P}\right)\left(\frac{\partial X}{\partial N_i}\right) \overline{\Delta P \Delta N_i}. \quad (19)$$

Fluctuations of the solvent are further characterized by

$$\overline{\Delta T^2} = \frac{kT^2}{C_p}, \quad \overline{\Delta P^2} = kT/2 \left(\frac{\partial V}{\partial P}\right)_{T,N_i},$$

$$\overline{\Delta N_i^2} = kT/\left(\frac{\partial \mu_i}{\partial N_i}\right)_{T,P}, \quad \overline{\Delta N_i \Delta N_j} = 0 \quad (i \neq j), \quad \overline{\Delta T \Delta P} = kT/2 \left(\frac{\partial V}{\partial T}\right)_{P,N_i}, \quad \overline{\Delta T \Delta N_i} = 0, \quad \overline{\Delta P \Delta N_i} = 0. \quad (20)$$

Local fluctuations in $\overline{\Delta T^2}$; $\overline{\Delta P^2}$ and $\overline{\Delta T \cdot \Delta P}$ are inversely proportional to the square root of the number of solvent molecules in the volume under consideration. On the other hand the fluctuation $\overline{\Delta N_i^2}$ is inversely proportional to the square root of the number of the $i$th component molecules in that volume. If the structural state of a protein depends critically on a low concentration of small molecules or ions, the structural fluctuation due to these may not be neglected.

### *8.2. Fluctuations of the structural state at constant solvent condition*

In the case of a "gradual structural change" where the probability distribution of structural state has only one maximum during the change, the average fluctuation of structural state $\overline{\Delta X^2}$ is inversely proportional to the curvature of the free energy plot at the minimum point,

$$\overline{\Delta X^2} = \overline{(X - X_m)^2} = \frac{\int_{-1}^{1} \Delta X^2 e^{-\Delta G/kT} dX}{\int_{-1}^{1} e^{-\Delta G/kT} dX} = \frac{1}{\left(\frac{\partial^2 G(X)}{\partial X^2}\right)_{X=X_m}}, \quad (21)$$

where

$$\Delta G = G(X) - G(X_m) = \frac{1}{2}\left(\frac{\partial^2 G(X)}{\partial X^2}\right)_{X=X_m} \Delta X^2. \quad (22)$$

Since

$$\left(\frac{\partial^2 G}{\partial X^2}\right)_{X=X_m} = \tfrac{1}{2} N_0 Z J \left(BT \frac{1}{1 - X_m^2} - 1\right), \quad (23)$$

the average fluctuation increases as the value of $X_m$ approaches zero, when $BT_d > 1$. In fact, for the "gradual structural change" where $BT_d \equiv A/C > 1$, the apparent width in the probability distribution increase as $X_m$ approaches zero as shown in fig. 9. If $BT_d \equiv A/C = 1$, the average fluctuation at transition point becomes infinite as is the case of the second order transition.

In the case of a "structural transition", as discussed in the previous section, each protein molecule fluctuates between two different states and large fluctuations

in the structural state are expected near the transition temperature.

*8.3. Fluctuations of the local structural state in a structural state*

In a structural state where $Y$ is fixed, a local structural state characterized by $Y_s$ might fluctuate because of an uneven distribution of the bonded state throughout the molecule. The probability distribution of the local structural state is given by eq. (14). Fluctuations of the local structural state correspond to a loss in the mixing entropy of bonds and a small gain in enthalpy due to cooperative interaction. As an example, let us consider a simple fluctuation of the local structural state that $Y_s$ takes zero while the fraction of bonded state in a whole molecule $Y$ has its most probable value $Y_m$. The entropy difference $\Delta S^f$ accompanied with the fluctuation can be expressed by

$$\Delta S^f = k \ln \frac{(N_0-N_s)!}{(N_0 Y_m)!\,(N_0-N_s-N_0 Y_m)!} - k \ln \frac{N_0!}{(N_0 Y_m)!\,(N_0-N_0 Y_m)!} . \quad (24)$$

If we assume that $N_s/N_0$ is small compared to $Y_m$ or $1-Y_m$, $\Delta S^f$ is given approximately by

$$\Delta S^f \doteqdot k N_s \ln(1-Y_m) . \quad (25)$$

Thus, the entropy difference is negative. Similarly, the enthalpy difference $\Delta H^f$ is given by

$$\Delta H^f = ZJ(N_0-N_s)\left(\frac{N_0}{N_0-N_s} Y_m\right)\left(1-\frac{N_0}{N_0-N_s} Y_m\right) - ZJN_0 Y_m(1-Y_m) \doteqdot -ZJN_s Y_m^2 \quad (26)$$

Since the free energy difference of the fluctuation $\Delta G^f$ is always positive, the entropy loss must overcome the enthalpy gain.

## 9. Effect of solvent and small molecules

Here we consider a multi-component system in which a protein molecule interacts with the solvent and other small components. The protein molecule is assumed to have $n_i$ equivalent binding sites with the same binding constant $K_i$ for the $i$th component. The solvent and the other components are denoted by suffix $i = 0$ and $i = 1, 2, \ldots, n$. Then the free energy of the system can be written by

$$G(T,P,a_i,X) = N_0\epsilon\tfrac{1}{2}(1-X) + N_0 ZJ\tfrac{1}{4}(1-X^2) + N_0 kT[\tfrac{1}{2}(1+X)\ln\tfrac{1}{2}(1+X) + \tfrac{1}{2}(1-X)\ln\tfrac{1}{2}(1-X)] - N_0\tfrac{1}{2}(1-X)\alpha T - \sum_{i=0}^{n_0} n_i(X)kT\ln[1+K_i(X,T,P)a_i] . \quad (27)$$

The first four terms are the same as in eq. (1) and the last one is the interaction term where $a_i$ is the activity of the $i$th component.

Generally $n_i$ is a function of the structural state $X$ and $K_i$ is a function of $X$ and temperature $T$ and pressure $P$. Therefore the average structural state of the protein molecule is affected by binding of the $i$th component through $n_i(X)$ and $K_i(X, T, P)$.

*9.1. Structural changes by nonspecific binding*

Water molecules and several denaturants such as urea, guanidine hydrochloride and LiCl are considered to interact with protein molecule in a nonspecific manner. For such "nonspecific binding", we propose that the binding constant $K$ does not depend on the structural state characterized by $Y$ or $X$ but the number of binding sites is proportional to the fraction of unbonded state $1-Y$ [$=\frac{1}{2}(1-X)$]. Then the free energy of a system in which only the $i$th component binds to a protein molecule in a nonspecific manner is given by

$$G(T,P,a_i,X) = N_0\epsilon\tfrac{1}{2}(1-X) + N_0 ZJ\tfrac{1}{4}(1-X^2) + N_0 kT[\tfrac{1}{2}(1+X)\ln\tfrac{1}{2}(1+X) + \tfrac{1}{2}(1-X)\ln\tfrac{1}{2}(1-X)] - N_0\tfrac{1}{2}(1-X)\alpha T - N_0\tfrac{1}{2}(1-X)\,l_i kT\ln[1+K_i(T,P)a_i], \quad (28)$$

where $l_i$ is the number of sites per secondary bond.

From the condition of minimum free energy

$$\frac{\partial G(T,P,a_i,X)}{\partial X} = 0 ,$$

the most probable value of $X$, $X_m$, is given by

$$\tanh^{-1} X_m \equiv \frac{1}{2}\ln\left(\frac{1+X_m}{1-X_m}\right) = \frac{A+X_m}{BT} - (C+D_i), \quad (29)$$

where $A, B, C$ were defined in eq. (3) and $D_i$ is given by

$$D_i(T, P, a_i) = \tfrac{1}{2} l_i \ln(1 + K_i(T, P) a_i) \qquad (30)$$

The only difference between eq. (29) and eq. (7) is that the former has the term $(C + D_i(T, P, a_i))$ instead of $C$ in the latter. The "characteristic point" moves a distance $D_i$ to the $-x$ direction from the point $p(-C, -A)$ with increasing concentration or activity of the $i$th component. The structural state changes always to a more disordered state by the addition of the $i$th component.

A transition or sharp change of the structural state is expected at a certain concentration of the $i$th component though the distance $D_i$ changes gradually according to the logarithmic function in eq. (30). As the concentration of the $i$th component increases, the most probable value, $X_m$ shown in fig. 14, changes gradually from the value corresponding to $Q_0$ to that corresponding to $Q_d$ and then changes discretely to the value corresponding to $S_d$ at a critical concentration of the $i$th component when $(T/T_d)(A/C) < 1$. The calculated values of the most probable and average values of $X$, $X_m$ and $\bar{X}$, are shown in fig. 15 as functions of $a_i$, for typical values of the parameters $A, B, C, N_0, K_i$ and temperature. Similar sharp transitions state have been reported in the denaturation process of many proteins by urea, guanidine hydrochloride (GuHCl) or other simple salts [2].

Usually heat-denatured proteins are not in a completely disordered state and further denaturation occurs when GuHCl is added after completion of the

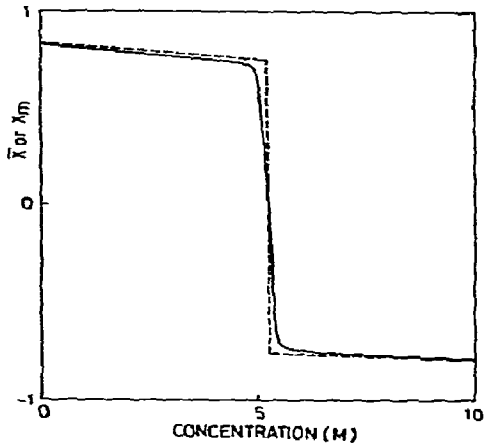


Fig. 15. The change in the average structural state $\bar{X}$ (solid line) and the most probable structural state $X_m$ (dotted line) with the concentration of a "nonspecific binding" component. The values of the parameters are $A = 0.5$, $B = 2.649 \times 10^{-3}$, $C = 0.5$, $N_0 = 500$, $T = 293°K$, $D(T) = 0.15\ ln(1 + K(T))$, $\ln K(T) = -8 + 2000/T$.

thermal denaturation [2]. Such cases are easily understood by the present model. If we assume a point $p_1$ in fig. 16 as the characteristic point of a protein molecule, the change in structural state with increasing temperature gives rise to a sharp transition from an ordered to a disordered state at the denaturation temperature $T_d$ although a considerable fraction of bonded state remains as indicated by point $S_d$ in fig. 16. By the

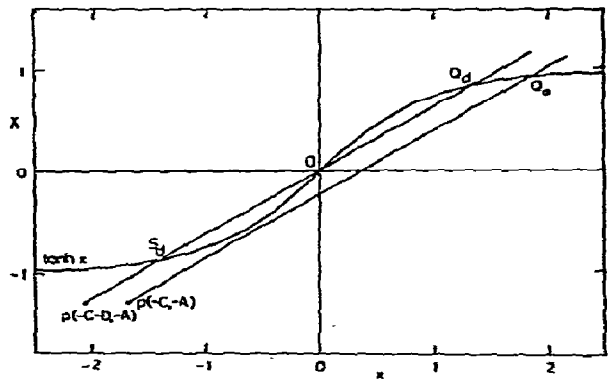


Fig. 14. The graphic analysis of the change in the most probable structural state $X_m$ with the concentration of a "nonspecific binding" component (see text).

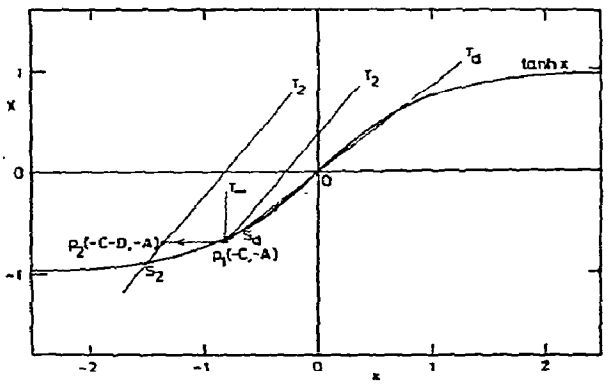


Fig. 16. The graphic analysis of the change in the most probable structural state $X_m$ induced by the addition of a denaturant after thermal denaturation (see text).

addition of denaturant at a certain temperature $T_2$, the characteristic point moves from $p_1$ to $p_2$ and the fraction of bonded state decreases even beyond the value corresponding to infinite temperature ($S_2$ in fig. 16).

Very interesting phenomena were reported for the structural change of β-Lactoglobulin and some other proteins induced by denaturants and temperature. As the temperature rises, the structural state of β-Lactoglobulin in the presence of a certain concentration of urea changes from the disordered state to the ordered state and again to the disordered state [13]. Such phenomena could simply be explained if we assume a proper function for the binding constant $K_i$ which decreases sharply with rising temperature. Since the nature of nonspecific binding of denaturants is supposed to be that of a simple physical adsorption or affinity rather than that of a chemical bond, such binding constants are expected to depend mainly on water structure and then on temperature.

Structural changes induced by denaturants and temperature are discussed in a general way considering the structure of water in the following paper.

### *9.2. Structural changes by specific binding*

Contrary to the case of nonspecific binding, we assume that the binding constant $K_j$ of specific molecules such as substrates, inhibitors and allosteric effectors depends on the structural state $X$ but the number of the binding sites is independent of $X$.

Then the free energy of a system in which the $j$th component binds specifically to the protein molecule is given by

$$G(T,P,a_j,X) = +N_0\epsilon\tfrac{1}{2}(1-X) + N_0ZJ\tfrac{1}{4}(1-X^2) + N_0kT[\tfrac{1}{2}(1+X)\ln\tfrac{1}{2}(1+X) + \tfrac{1}{2}(1-X)\ln\tfrac{1}{2}(1-X)] - N_0\tfrac{1}{2}(1-X)\alpha T - n_jkT\ln(1+K(X,T,P)a_i) . \quad (31)$$

From the condition of minimum free energy, it follows that

$$\tanh^{-1}X_m \equiv \frac{1}{2}\ln\left(\frac{1+X_m}{1-X_m}\right) = (A+X_m)/BT - (C-E_j) ,$$

$$E_j \equiv \frac{n_j}{N_0}\frac{a_j(\partial K_j/\partial X)_{X=X_m}}{1+K_j(X,T,P)a_j} . \quad (32)$$

Although the characteristic point of the protein is modified by the presence of the $j$th component, the modification is complicated for the shift itself is a function of the structural state $X$ as shown by eq. (32). The direction of modification of the characteristic points and then the direction of the structural change, however, can be summarized as follows:

(a) If the binding constant is independent of $X$, that is $\partial K_j/\partial X = 0$, no structural change occurs by the addition of the $j$th component.

(b) If the binding constant of the $j$th component increases with the increase of the structural parameter $X$, that is $\partial X_j/\partial X > 0$, the characteristic point moves to the $+x$ direction and the structural state changes to the more ordered state.

(c) If the binding constant of the $j$th component decreases with the increase of the structural parameter $X$, that is $\partial K_j/\partial X < 0$, the characteristic point moves to the $-x$ direction and the structural state changes to the direction of more disordered state.

In the case of (c), a drastic structural change could be expected at a certain concentration of the $j$th component as was the case for the nonspecific binding.

### *9.3. Allosteric effect and allosteric proteins*

The essential character of an allosteric effect is that the specific binding of a low molecular weight component or ligand to a protein molecule affects the affinity of another component to a sterically different site of the protein molecule.

In the present model, such an allosteric effect could be considered to arise from the general nature of protein and would be expected without any further assumption.

For the case of specific binding, the binding constant depends on the average structural state $X$ and the binding changes the structural state itself; therefore, the binding of a specific component is affected by the binding of other components through the modification of structural state $X$.

If a binding constant of a substrate s of an allosteric protein increases with the increase of structural parameter $X$, i.e. $\partial K_s/\partial X > 0$, all the specific components $i$ which have a similar binding nature, $\partial K_i/\partial X > 0$, might be allosteric inducers and all the other specific components $j$ which are of an opposite binding nature, $\partial K_j/\partial X < 0$, might be allosteric inhibitors. Thus, an allosteric protein has not necessarily to be an oligomer

but can be a monomer in as far as the heterotropic effect is concerned.

Since the homotropic effect, which assumes several identical binding sites in a protein, is one of the essential features of allosteric proteins, a so called allosteric protein should be an oligomer. Positive or negative homotropic effect could be explained in connection to the change of the structural state of subunits associated with the formation of the oligomer. Details will be discussed in following papers.

## 10. Discussion

The present theory based on the statistical thermodynamic model at the chain level has many advantages over those proposed previously to explain the conformational change of proteins.

The theory reveals the thermal denaturation in detail, such as the temperature dependence of the probability distribution and the heat capacity. It predicts either a "structural transition" or a "gradual structural change" depending on the position of the "characteristic point" of the protein molecule. It also predicts a small reconversion with a further rise in temperature after the sharp change from the ordered to the disordered state in some "structural transitions".

The so-called two state model usually accepted to analyze thermal denaturation of proteins is contained in the theory as a special case of the "structural transition".

Poland and Sheraga [4] pointed out the simplicity of the two state model and they proposed a conceptual multi-state model. They predicted two alternative temperature dependences for the relative probability of conformational species. In one case the most probable species changes continuously from the "native" to the "denative" state with rise in temperature. This prediction corresponds to a "gradual structural change" in our theory. In the other prediction, the relative probabilities for all species are equal at the transition temperature. This corresponds to the special case, $A = C$, in our model where the fluctuation of the structural state become infinite as discussed in the previous section.

One of the important predictions of the theory is that so-called "native state" in solution at room temperature is not a fully fixed structure but a fairly versatile one. A considerable fraction of secondary bonds are in the unbonded state on average. The number and position of the secondary bonds in the bonded state are always fluctuating. Quantitative expressions of the structural fluctuations are given in connection with the equilibrium properties. Such a fluctuating model might be useful for understanding the dynamical properties of protein molecules.

In the present theory, structural changes induced by the specific binding of low molecular components is discussed in a unified manner. Koshland [14] proposed that protein molecules change their conformation to a more suitable one for the binding when they interact with substrates. This "induced fit" model could correspond to the case where the binding constant of the substrate increases with the structural parameter $X$, $\partial K_s/\partial X > 0$. Thus, the interaction between protein and substrate increases ordered structures and reduces possible structures or fluctuations.

In the "lack model" proposed by Lumry [15], the structure of the substrate was assumed to be distorted by the binding to the enzyme. The model may correspond to the case that $\partial K_s/\partial X < 0$, where the binding of the substrate is favorable in the more disordered state of the enzyme.

The most characteristic aspect of the theory is that the allosteric effect can be explained by the general structural change induced by a specific binding without any further assumption.

An important advantage of the theory is that the value of the parameters can be, in principle, determined from the nature of the bonds, the primary structure and the tertiary structure at 0°K or in the crystal. Therefore, the theory is the first step for the prediction of the structural changes of a protein from the knowledge of the crystal structure. In fact, the value of the parameters determined from the experimental results for several grobular proteins reflect the molecular properties of each protein well. Details will be shown in the following paper.

Although the essential factors describing the molecular nature of proteins are included in the model, the assumptions used in the formulation of the free-energy are oversimplifications. A more precise treatment of the present model might be necessary for making predictions. Considerations about the dissimilar nature of the secondary bonds or chains, the effect of the lattice surface and a more precise expression of the

entropy or cooperative actions are required for further studies.

## Acknowledgement

The author wishes to thank Dr. M.I. Kanehisa for computer calculations.

## References

[1] W. Kauzmann, Advanc. Protein Chem. 14 (1959) 1.
[2] C. Tanford, Advanc. Protein Chem. 23 (1968) 121.
[3] R. Lumry, R. Biltonen and J.F. Brandts, Biopolymers 4 (1966) 917.
[4] D. Poland and H.A. Scheraga, poly-αAmino acid, ed. G.D. Fasman (Marcel Dekker, New York, N.Y., 1967) p. 391.
[5] J.F. Brandts, Structure and Stability of Biological Macromolecules, ed. S.N. Timasheff and G.D. Fasman (Marcel Dekker, New York, N.Y., 1969) p. 213.
[6] J. Monod, J. Wyman and J.P. Changeux, J. Mol. Biol. 12 (1965) 88.
[7] D.E. Koshland, Jr., G. Némethy and D. Filmer, Biochemistry 5 (1966) 365.
[8] M. Nakanishi, M. Tsuboi and A. Ikegami, J. Mol. Biol. 75 (1973) 673.
[9] S.W. Englander and A. Rolfe, J. Biol. Chem. 248 (1973) 4852.
[10] J. Hermans, Jr., D. Lohr and D. Ferro, Advanc. Polym. Sci. 9 (1972) 230.
[11] J.F. Brandts and L. Hunt, J. Am. Chem. Soc. 89 (1967) 4826.
[12] P.L. Privalov, N.N. Khechinashvili and B.P. Atanasov, Biopolymers 10 (1971) 1865.
[13] N.C. Pace and C. Tanford, Biochemistry 7 (1968) 198.
[14] D.E. Koshland, Jr., The Enzymes, 2nd ed., ed. Boyer et al. 1 (1959) 305.
[15] R. Lumry, The Enzymes, 2nd ed., ed. Boyer et al. 1 (1959) 157.