Biophysical Chemistry 6 (1977) 131–149

STRUCTURAL CHANGES AND FLUCTUATIONS OF PROTEINS.
II. Analysis of the denaturation of globular proteins

Minoru I. KANEHISA * and Akira IKEGAMI
Department of Physics, Faculty of Science, University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113, Japan

Received 3 August 1976

The statistical thermodynamic model of protein structure proposed in paper I is developed with special attention to the hydrophobic interaction. Calorimetric measurements of the thermal denaturation of five globular proteins, ribonuclease A, lysozyme, α-chymotrypsin, cytochrome c, and myoglobin, are quantitatively analyzed using the model. The thermodynamic parameters obtained by the least squares method reflect the global, average properties of proteins and are in good agreement with the expected values estimated from experimental and theoretical studies for model peptides. The average bond energy ϵ is well related to the tertiary structure of each protein. However, the difference in the parameters between different proteins is not observed for the cooperative energy ZJ and the chain entropy α. The individuality of a protein as far as its structural stability is concerned, is mainly reflected by the parameter γ specifying the hydrophobic nature of a protein. The model is further applied in the analysis of several aspects of the structural stability of globular proteins. Denaturation induced by denaturants, salts, and pH are also explained by the model in a unified manner.

1. Introduction

The statistical thermodynamic model of the structure of proteins proposed in the preceding paper (I) is applied in the analysis of the actual denaturation of globular proteins. This model has the following distinct features. The "structure" of a protein is characterized by the bonding state of a set of "identical" secondary bonds which are "uniformly" located at lattice points in the globule. The macroscopic state of a protein specified by the fraction of secondary bonds in the bonded state is called the "structural state" in which a number of structures are included. As a consequence of the "cooperativity" between bonds, a protein molecule often undergoes the "two-state like structural transition", in which two structural states corresponding to the native and denatured state may be distinguished.

The recent appearance of a new measuring technique – differential scanning calorimetry – has made it possible to obtain the heat capacity of a protein in dilute solutions with sufficient precision and reproducibility. Calorimetric data, which are the direct reflection of the nonspecific, average properties of a protein as a whole, are open to analysis by a proper method other than the traditional two-state analysis [1–3]. Our model provides quite a suitable treatment of such data to estimate thermodynamic parameters of globular proteins.

Since Kauzmann [4] discussed the thermodynamic characteristics of the so-called "hydrophobic bond" in relation to the "iceberg" concept for clathrated hydrocarbons in water due to Frank and Evans [5], it has been widely recognized that the properties of proteins in aqueous solutions must be rationalized on the basis of solvent structural effects. However, the model described in the preceding paper did not include such considerations and some quantitative discrepancies are observed; the specific heat always falls to zero after the transition in the preceding treatment (see fig. 11 of the preceding paper) while experimental results show that the denatured state of any protein

* Present address: Department of Physiological Chemistry, The Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA.

has a much higher value than the native one on account of the exposure of the nonpolar groups to the solvent (see fig. 2b). Moreover, sometimes in the presence of denaturants the parabolic-type transition (see fig. 8a) is observed with the rise of temperature, which can also be explained by solvent structural effects. Therefore, we shall first introduce this effect of water structure into our model of the protein structure.

2. Free energy of protein in aqueous solutions

Frank and Evans [5] postulated the formation of hydrogen-bonded quasi-crystalline structures of water molecules (icebergs) around isolated, nonpolar molecules in aqueous solution. This hypothesis successfully explains, at least qualitatively, the anomalous thermodynamic behaviour associated with the transfer of a nonpolar solute from a hydrocarbon environment to water, that is, the unusual negative enthalpy change ($\Delta H < 0$) and very large negative excess entropy change ($\Delta S < 0$) over the entropy of ideal mixing, which leads to a large positive free energy change ($\Delta G > 0$) and hence to a low solubility. The melting of these ice-like structures with increasing temperature also accounts for the much higher partial molal heat capacity ($\Delta C_p > 0$) of nonpolar solutes in aqueous solution than in the pure liquid. The change in the partial molal volume ($\Delta V < 0$) is also explained by the decrease in the coordination number of the water molecule.

One common feature of the structure of globular proteins, as revealed by X-ray crystallography [6], is that the polar residues are situated on the surface and that most of the nonpolar side chains are buried in the interior. A hydrophobic bond in a protein is considered to be formed when nonpolar side chains tend to adhere so as to reduce unfavourable contacts with the surrounding water [4,7]. It is not primarily maintained by van der Waals interactions, but rather by entropic effects of the solvent. In this sense it is not an actual energetic bond, but a statistical thermodynamic bond.

Actually the hydrophobic bonding has an anomalous thermodynamic behaviour and hence it must be taken into account separately in our free energy expression. Since our model describes only the changes between native and denatured states, the surface, which is already in contact with water in the native state, is not considered here. On the other hand nonpolar side chains buried in the interior bring about changes in the surrounding water structure when they establish new contacts with water upon unfolding of the molecule. Thus, we include the following additional description in our model.

When one secondary bond changes from the bonded to the unbonded state, γ water molecules are brought into new contact with nonpolar groups. The free energy difference of the water molecule between the bulk water and the first layer around a nonpolar group is Δg_w.

Note that the contributions to hydrophobic bonding from Van der Waals interactions and internal bond rotations are included in our model through the bond energy ϵ and the chain entropy α, respectively, so that the term Δg_w represents only changes in the water structure.

The explicit form of Δg_w is borrowed from the theoretical computation of Némethy and Scheraga [8], who gave quantitative estimates for the thermodynamic parameters of hydrophobic bonds on the basis of their statistical mechanical treatments of pure water [9] and of aqueous hydrocarbon solution [10]. According to their calculation, the free energy difference per mole of water undergoing a change in state from the bulk of pure liquid to the first layer next to a nonpolar substance is expressed as

$$\Delta g_w = a + bT + cT^2 \,. \tag{1}$$

Some numerical values are

$$\Delta g_w = -957 + 6.08T - 0.00824T^2 \,, \tag{2}$$

in cal/mole for aliphatic side chains and

$$\Delta g_w = -703 + 4.35T - 0.00617T^2 \,, \tag{3}$$

in cal/mole for aromatic side chains [8]. Their calculations are in fairly good agreement with experimental data of the thermodynamic parameters for the transfer of hydrocarbons [10]. Hence, we make use of eqs. (2) and (3) in determining the thermodynamic parameters of globular proteins.

Then, the Gibbs free energy per mole of protein in aqueous solution takes the form (cf., eq. (1) of I)

$$\begin{aligned} G(T,X) = {} & N_0\{\epsilon - \alpha T + \gamma(a + bT + cT^2)\}\tfrac{1}{2}(1-X) \\ & + N_0 Z J \tfrac{1}{4}(1-X^2) + N_0 RT[\tfrac{1}{2}(1+X)\ln\tfrac{1}{2}(1+X) \\ & + \tfrac{1}{2}(1-X)\ln\tfrac{1}{2}(1-X)] \,, \end{aligned} \tag{4}$$

or in a more compact form

$$G(T, X) = N_0(\epsilon' - \alpha' T - \beta' T^2)\tfrac{1}{2}(1-X) + N_0 ZJ\tfrac{1}{4}(1-X^2)$$
$$+ N_0 RT[\tfrac{1}{2}(1+X)\ln\tfrac{1}{2}(1+X) + \tfrac{1}{2}(1-X)\ln\tfrac{1}{2}(1-X)], \quad (5)$$

where

$$\epsilon' = \epsilon + \gamma a , \quad (6)$$

$$\alpha' = \alpha - \gamma b , \quad (7)$$

$$\beta' = -\gamma c . \quad (8)$$

From this free energy, the temperature dependent enthalpy and entropy are calculated

$$H(T, X) = N_0(\epsilon' + \beta' T^2)\tfrac{1}{2}(1-X) + N_0 ZJ\tfrac{1}{4}(1-X^2), \quad (9)$$

$$S(T, X) = N_0(\alpha' + 2\beta' T)\tfrac{1}{2}(1-X)$$
$$- N_0 R[\tfrac{1}{2}(1+X)\ln\tfrac{1}{2}(1+X) + \tfrac{1}{2}(1-X)\ln\tfrac{1}{2}(1-X)] . \quad (10)$$

Average physical quantities are calculated in the same manner as in I. The average enthalpy at temperature T is, from eq. (9),

$$\bar{H}(T) = N_0(\epsilon' + \beta' T^2)\tfrac{1}{2}(1-\bar{X}) + N_0 ZJ\tfrac{1}{4}(1-\overline{X^2}) , \quad (11)$$

therefore, the heat capacity is

$$\bar{C}_p(T) = -\tfrac{1}{2}N_0(\epsilon' + \beta' T^2)\frac{\partial \bar{X}}{\partial T} -$$
$$- \tfrac{1}{4}N_0 ZJ \frac{\partial \overline{X^2}}{\partial T} + N_0\beta' T(1-\bar{X})$$
$$= \frac{N_0^2}{4RT^2}[\epsilon'^2(\overline{X^2} - (\bar{X})^2) + \epsilon' ZJ(\overline{X^3} - \bar{X}\,\overline{X^2})$$
$$+ \tfrac{1}{4}(ZJ)^2(\overline{X^4} - (\overline{X^2})^2)] + N_0\beta' T(1-\bar{X}). \quad (12)$$

The thermal denaturation of proteins can be discussed through the temperature dependence of the thermodynamic functions, $\bar{X}$, $\bar{H}$, and $\bar{C}_p$. Sample calculations of the structural state or the order parameter $\bar{X}$ and the specific heat $\bar{C}_p$ are shown as solid lines in fig. 1a and b. Dotted lines represent the corresponding calculations without taking into consideration the water structure. Since eq. (5) can be written

$$G(T, X) = N_0(h(T) - S(T)T)\tfrac{1}{2}(1-X) + N_0 ZJ\tfrac{1}{4}(1-X^2)$$
$$+ N_0 RT[\tfrac{1}{2}(1+X)\ln\tfrac{1}{2}(1+X) + \tfrac{1}{2}(1-X)\ln\tfrac{1}{2}(1-X)], \quad (13)$$

where

$$h(T) = \epsilon' + \beta' T^2 , \quad (14)$$

$$s(T) = \alpha' + 2\beta' T , \quad (15)$$

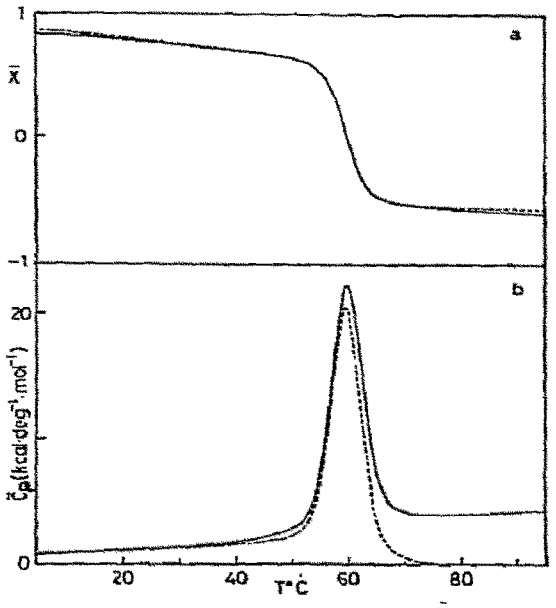

Fig. 1. Comparison of thermal denaturation curves with and without consideration of the water structure; (a) structural state $\bar{X}$; (b) specific heat $\bar{C}_p$. Solid lines are calculated from the free energy of eq. (13), with $N_0 = 250$, $T_d = 60°C$, $\epsilon' = -2$ kcal/mole, $\beta' T_d^2 = 3$ kcal/mole, and $ZJ = 1.5$ kcal/mole (hence, $h(T_d) = \epsilon' + \beta' T_d^2 = 1$ kcal/mole). Dotted lines are corresponding curves without consideration of the water structure calculated from the free energy of eq. (4) of paper I, with $N_0 = 250$, $T_d = 60°C$, $\epsilon = 1$ kcal/mole, and $ZJ = 1.5$ kcal/mole.

the corresponding parameters of eq. (1) in paper I are chosen as $\epsilon = h(T_d)$ and $\alpha = s(T_d)$. As is illustrated in these figures the thermal denaturation as observed by the order parameter can be treated as if the protein virtually in a vacuum, whereas the effect of the water structure must be introduced separately so as to explain the higher value of the specific heat in the denatured state than in the native state. Note that the thermodynamic parameters ϵ and α used in the paper must be regarded as an "effective" bond energy and chain entropy containing the contribution from the water structure.

3. Least squares fit of experimental data

The thermodynamic parameters of our model, N_0, ϵ, α, ZJ, and γ, are determined for several globular

proteins from precise calorimetric measurements of thermal denaturation in aqueous solutions. Since calorimetric data directly reflect nonspecific, average thermodynamic properties of a protein as a whole, they are directly related to our thermodynamic functions, $\bar{H}$ and $\bar{C}_p$, so long as the "identical and uniform" assumption for the secondary bonds is valid. One thing to note is that it is only the thermodynamic behaviour during structural transitions that are treated in our model. To exclude the contributions from the internal degrees of freedom to the absolute specific heat, they have to be evaluated separately (see appendix).

Provided that the denaturation temperature T_d is known experimentally, the parameters to be determined are reduced to four, N_0, ϵ', β', and ZJ, since eq. (5) is transformed as

$$G(T, X) = N_0\epsilon'\left(1 - \frac{T}{T_d}\right)\tfrac{1}{2}(1-X) + N_0 ZJ\tfrac{1}{4}(1-X^2) + N_0\beta' T_d^2\left(\frac{T}{T_d} - \frac{T^2}{T_d^2}\right)\tfrac{1}{2}(1-X) + N_0 RT[\tfrac{1}{2}(1+X)\ln\tfrac{1}{2}(1+X) + \tfrac{1}{2}(1-X)\ln\tfrac{1}{2}(1-X)] \quad (16)$$

These parameters are determined by the least squares method. The optimization function is

$$I(P) = \sum_{i=1}^{\text{No. of data}} [\bar{S}(P; T_i) - S_i^{\exp}]^2 \quad (17)$$

where P represents symbolically the parameters to be fitted and the superscript exp denotes an experimental value. The physical quantity S may be the enthalpy H or specific heat C_p and its calculated value $\bar{S}$ is given by eq. (11) or (12). In practice the parameter N_0 is not determined directly in our optimization routine. This routine, which determines the value of the three energies ϵ', $\beta' T_d^2$, and ZJ for each run, operates for continuously changing values of N_0. The optimized point (local minimum of the function I) is searched by the variable metric method of Fletcher and Powell [11].

In order to determine the parameters promptly and definitively, we have designed a man–machine interactive system by the graphic programming for this nonlinear least squares fit. Figs. 2a and b show pictures on the screen of HITAC 8811 Display Units attached to HITAC 8800/8700 Computing System at Tokyo

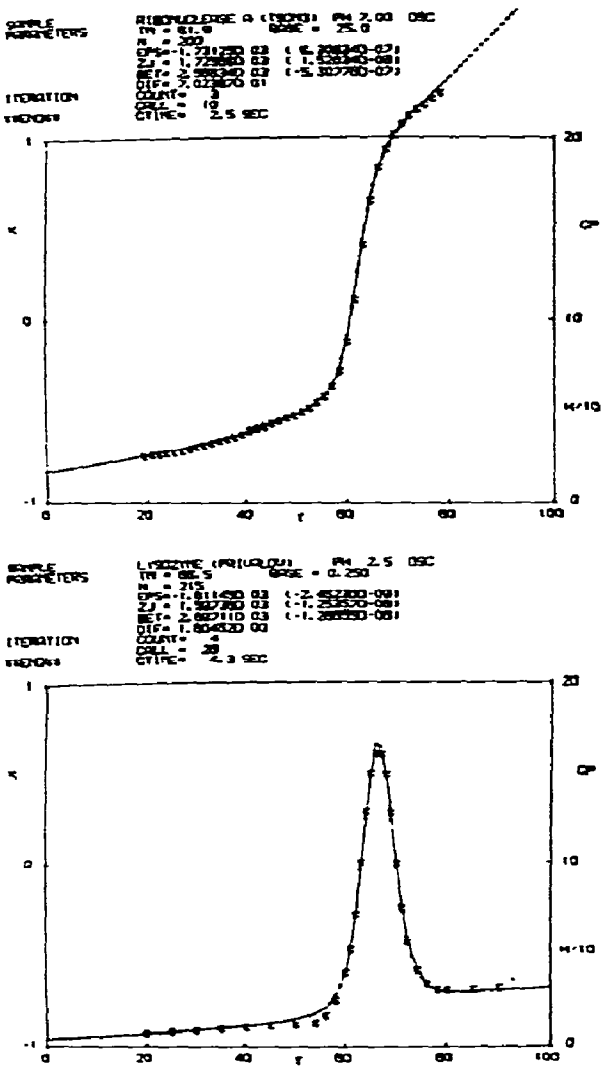

Fig. 2. Pictures on the graphic display in our optimization routine; * indicates an experimental datum and the solid line is the optimized curve; (a) the enthalpy of ribonuclease A [12]; (b) the specific heat of lysozyme [13].

University Computer Center. The calculated curve is displayed together with experimental values at every moment in iterations so as to visualize whether this routine is really optimizing and the minimization can be easily restarted at different initial points to check that the minimum point is unique.

The experimental results analyzed by our method

are those by Tsong et al. [12] and by Privalov and Khechinashvili [13]. The fitting procedure is as follows.

3.1. Calorimetric measurements of ribonuclease A and its derivatives

The thermal transitions of bovine pancreatic ribonuclease A and certain of its derivatives have been studied by Tsong et al. [12] in a highly sensitive differential scanning calorimeter, which records the difference in power necessary to heat equally two identical cells, one with the protein solution and the other with the solvent. The primary data obtained by them consisted in values of the total energy fed back to the calorimeter and the temperature at intervals of 1 min. These data are converted to the excess enthalpies of the protein solution relative to the solution at 0–5°C, where the instrument is adjusted to give a horizontal base line (fig. 2a).

This enthalpy, of which numerical values are tabulated in their article, is directly related to our thermodynamic function $\bar{H}$, if its absolute value in the low temperature region is known. However, as we have no information about this value, it is also treated as a parameter to be adjusted in our optimization routine. We have obtained the value of $H(0°C)$ as 25 kcal/mole which seems to give the best fit to the experimental results at all pH values. Although a change of this value may cause some modification of the absolute value of the other parameters (for example, using the value of 20 kcal/mole causes the difference in ϵ', $\beta' T_d^2$, or ZJ not to exceed 0.1 kcal/mole), the overall features of the pH dependence are not varied. The denaturation temperature T_d has been determined to be at a slightly higher value (less than 1°C) than the one obtained by the authors on the basis of a two-state analysis.

Thus, the three parameters, ϵ', $\beta' T_d^2$, and ZJ, have been determined by the least squares method with continuously varying values of N_0. Table 1 shows the dependence on N_0. We have assumed the value of 200 in obtaining the results of table 2, although this value is rather ambiguous. This point will be discussed later.

3.2. Calorimetric measurements of five globular proteins

Recently Privalov and Khechinashvili [13] have investigated the thermal properties of five globular proteins with known three-dimensional structures, ribonuclease A, lysozyme, α-chymotrypsin, cytochrome c, and myoglobin, by their new model of a differential

Table 1
Calculated thermodynamic parameters for ribonuclease with various values of N [a]

Protein	pH	T_d	N_0	ϵ'	$\beta' T_d^2$	ZJ	ϵ	α	γ	Variance
Ribonuclease A	0.36	32.5	100	−2.18	3.04	1.54	1.59	6.85	4.03	2.50
			150	−1.37	1.87	1.67	0.960	4.15	2.49	0.962
			200	−0.922	1.28	1.80	0.667	2.87	1.70	0.568
			250	−0.690	0.967	1.93	0.510	2.19	1.29	0.475
			300	−0.562	0.783	2.03	0.410	1.76	1.04	0.881
	3.28	46.3	100	−2.62	3.94	1.55	1.85	7.84	4.79	9.02
			150	−1.96	2.72	1.65	1.14	4.98	3.31	2.37
			200	−1.38	1.91	1.77	0.793	3.48	2.33	1.31
			250	−1.03	1.44	1.89	0.603	2.63	1.75	1.67
			300	−0.808	1.14	2.01	0.487	2.12	1.39	1.63
	7.80	61.9	100	−2.14	4.29	1.51	2.29	8.93	4.75	28.3
			150	−2.42	3.67	1.58	1.37	5.86	4.06	8.28
			200	−2.13	2.96	1.65	0.927	4.20	3.27	2.41
			250	−1.68	2.29	1.75	0.690	3.17	2.54	3.98
			300	−1.34	1.83	1.84	0.548	2.52	2.02	7.91

[a] In tables 1, 2, and 3, the energy is expressed in kcal/mole, the entrope in eu, and T_d in °C. The variance is relative to the value of eq. (17) at the local minimum divided by the number of data points.

Table 2
Calculated thermodynamic parameters for ribonuclease at various values of pH

Protein	pH	T_d	N_0	ϵ'	$\beta' T_d^2$	ZJ	ϵ	α	γ	Variance
Ribonuclease A	0.36	32.5	200	−0.922	1.28	1.80	0.667	2.87	1.70	0.568
	1.05	30.8		−1.01	1.36	1.79	0.691	3.01	1.82	0.326
	2.02	33.8		−0.986	1.36	1.85	0.694	2.99	1.80	1.14
	2.80	41.7		−1.27	1.73	1.78	0.756	3.32	2.17	0.938
	3.28	46.3		−1.38	1.91	1.77	0.793	3.48	2.33	1.31
	4.04	52.8		−2.07	2.66	1.70	0.828	3.89	3.11	3.96
	5.00	58.3		−2.53	3.20	1.64	0.854	4.16	3.62	6.30
	6.23	61.4		−2.05	2.78	1.71	0.832	3.84	3.08	2.74
	7.00	61.9		−1.73	2.57	1.73	0.920	3.99	2.84	1.76
	7.80	61.9		−2.13	2.96	1.65	0.927	4.20	3.27	2.41
Ribonuclease S	7.00	48.8	200	−1.26	1.81	1.76	0.770	3.32	2.17	2.08
S protein	7.00	38.3	168	−1.17	1.51	1.60	0.640	2.86	1.93	0.580

scanning calorimeter. The difference between two calorimetric records, one of which is the baseline obtained by filling two cells with solvent and the other is measured after filling one of the cells with the protein solution, was converted to the partial specific heat of protein in dilute solution such as shown in fig. 2b. Although the baseline stability still is a subtle factor, the partial specific heats of all the proteins studied are reported to be the same within experimental error; at 20°C they are equal to $C_p(20°C) = 0.31 \pm 0.02$ cal $g^{-1} deg^{-1}$. Moreover, the partial specific heat increases linearly with temperature in both the native and denatured state, where the intensive heat absorption connected with the structural transition is absent, and its temperature dependence seems to be nearly equal for all the proteins studied, being $dC_p/dT = (2.2 \pm 0.5) \times 10^{-3}$ cal $g^{-1} deg^{-2}$.

As mentioned above such absolute values contain contributions from the internal degrees of freedom which are not included in our model. These contributions are estimated in the appendix and the results for these five globular proteins are nearly the same; $C_p(20°C) = 0.25$ cal $g^{-1} deg^{-1}$ and $dC_p/dT = 1 \times 10^{-3}$ cal $g^{-1} deg^{-2}$.

After subtraction of these values from those obtained by the authors, the three parameters, ϵ', $\beta' T_d^2$, and ZJ, have determined by least squares with the values of N_0 listed in table 3. These N_0 values are tentatively adopted by assuming that all the proteins contain the same number of secondary bonds per gram, hence N_0 is proportional to the molecular weight (table 5), and that the value for ribonuclease A is around 200.

3.3. Separation of the contribution from the water structure

To obtain the thermodynamic parameters of a protein itself, the contribution from the changes in the water structure must be separated according to eqs. (6), (7), and (8). The parameters a, b, and c of eq. (4) are determined for each protein from its amino acid composition, averaging eqs. (2) and (3), and shown in table 4. The values of ϵ, α, and γ thus obtained are added to the results of tables 1, 2, and 3.

4. Thermodynamic parameters of proteins

On the basis of the results shown in tables 1 and 3, the thermodynamic parameters of these globular proteins are here compared with former estimates and moreover analyzed in detail with reference to the molecular parameters determined from their primary and tertiary structure. Privalov and Khechinashvili [13] have already shown in their original article that the increase in the specific heat in the denatured state observed by them is satisfactorily related to the number of nonpolar contacts computed from the atomic coordinates determined by X-ray crystallographers. Their computation is shown in table 5, as well as other structural parameters for the five proteins. The

Table 3
Calculated thermodynamic parameters for five globular proteins

Protein	pH	T_d	N_0	ϵ'	$\beta' T_d^2$	ZJ	ϵ	α	γ	Variance
Myoglobin	12.0	54	269	−2.61	3.19	1.45	0.846	4.18	3.73	2.72
	11.5	63		−2.55	3.26	1.49	0.789	3.90	3.61	4.97
	11.1	70.5		−2.41	3.26	1.54	0.788	3.78	3.46	5.16
	10.7	76		−3.21	4.14	1.56	0.724	3.89	4.24	2.45
Cytochrome c	3.0	59.5	186	−2.43	3.14	1.48	0.851	4.06	3.58	2.64
	3.7	70		−2.84	3.77	1.57	0.872	4.24	4.04	4.08
	4.6	77.5		−2.47	3.40	1.66	0.733	3.55	3.49	3.00
α-Chymotrypsin	2.6	44	378	−1.65	2.04	1.76	0.702	3.29	2.51	6.05
	3.0	52		−1.99	2.55	1.56	0.812	3.80	2.99	3.43
	4.0	57.5		−2.33	3.11	1.48	0.965	4.46	3.52	9.19
Lysozyme	2.0	56	215	−1.68	2.40	1.54	0.898	3.93	2.75	2.36
	2.5	66.5		−1.81	2.70	1.60	0.901	3.93	2.90	3.04
	4.5	78.5		−1.96	3.00	1.65	0.854	3.75	3.01	13.3
Ribonuclease A	2.7	43	204	−1.07	1.70	1.50	0.904	3.75	2.11	10.1
	3.7	55		−1.66	2.49	1.55	1.02	4.35	2.87	4.10
	5.4	63.5		−1.54	2.42	1.62	0.937	3.94	2.65	4.53

Table 4
Nonpolar amino acids composition and the parameters a, b, and c [a]

Protein	Source	D.P.	Ala (8)	Val (12)	Leu (13)	Ile (13)	Met (15)	Cys (10)	Pro (12)	Phe (15)	$a \times 10^{-2}$	b	$c \times 10^{3}$	$\frac{w_Q}{w_Q+w_r}$
Cytochrome c	Bovine heart	104	6	3	6	6	2	2	4	4	−9.19	5.82	−7.39	0.849
Ribonuclease	Bovine pancreas	124	12	9	2	3	4	8	4	3	−9.34	5.92	−8.05	0.910
Lysozyme	Hen egg while	129	12	6	8	6	2	8	2	3	−9.35	5.93	−8.06	0.915
Myoglobin	Sperm whale	153	17	8	18	9	2	0	4	6	−9.27	5.87	−7.99	0.880
β-Lactoglobulin	Bovine milk	162	14	10	22	10	4	5	8	4	−9.40	5.97	−8.10	0.934
α-Chymotrypsin	Bovine pancreas	241	22	23	19	10	2	10	9	6	−9.37	5.95	−8.08	0.922
Flagellin	Salmonella SJ 25	541	84	30	44	25	3	0	5	8	−9.43	5.98	−8.12	0.944

[a] Parameters a, b, and c are expressed so as to give Δg_w in cal/mole, for use in eq. (4). They are determined as a weighted average of eqs. (2) and (3), assuming the numbers of water molecules which can coordinate around these aliphatic and aromatic side chains, denoted as w_Q and w_r, respectively. Hence, $w_Q/(w_Q+w_r)$ is called the relative aliphatic content. The numerical value for each amino acid is taken from Némethy and Scheraga [8] and shown in parentheses.

number of polar contacts is defined as the number of pairs of polar groups located at a distance of 3.2 Å. Nonpolar contacts are defined as the couples of nonpolar groups, such as CH_3, CH_2, and CH, situated at a distance within 4.0 to 4.5 Å.

4.1. Number of secondary bonds

First, let us discuss the value of N_0. As shown in table 1, the experimental data of ribonuclease A by Tsong et al. [12] can be fitted with various values of N_0, and we have concluded that the value of 200 seems to be most suitable to reproduce their experiments for all pH values. However, their data can be also analyzed using the value of 180, for example, to give qualitatively similar results. On account of the high cooperativity between secondary bonds, the number of bonds N_0 is rather ambiguous. If we take two actual secondary bonds, instead of one, as a unit, the values of ϵ, α, and γ may be roughly doubled

Table 5
Structural parameters of five globular proteins [a]

	Mgl	Cyr	Chm	Lys	Rna
Molecular weight	17,900	12,400	25,200	14,300	13,600
Number of residues	153+heme	104+heme	241	129	124
Number of difulfide bonds	–	–	5	4	4
Number of polar contacts (b–b) [13]	98	37	105	54	53
(b–s)	22 133	23 70	54 173	20 89	19 81
(s–s)	13	10	14	15	9
Number of nonpolar contacts (4.0 Å) [13]	213	136	238	108	90
(4.5 Å)	456	298	586	302	209
Percentage of α-helix content [26]	65–77		12–13	28–42	6–18
β-structure content	0		23–32	10	36
random coil content	32–23		66–55	62–48	58–46
Number of single bonds about which rotation can occur [b]	725	487	1053	578	563
Isoelectric point [27]	(6.7) [c]	10.7	8.1	10.7	9.5
Partial specific volume [28]	0.743	0.72	0.721	0.703	0.692
Molecular dimension [28]	44×44×25	25×25×37		45×30×30	

[a] Abbreviations: Mgl, myoglobin; Cyr, cytochrome c; Chm, α-chymotrypsin; Lys, lysozyme; Rna, ribonucleas A; b–b, backbone–backbone; b–s, backbone–side chain; s–s, side chain–side chain.
[b] Our computation.
[c] Datum for hemoglobin.

(table 1). This would be often the case of hydrophobic bonds, where more than one pairs of van der Waals contacts can be formed and broken at a time. Nevertheless, our estimation of N_0 seems to be fairly good in relation to the computed bonds as shown in table 6, where the numbers per N_0 of polar contacts (hydrogen bonds) and nonpolar contacts within 4.0 Å (hydrophobic bonds) are presented. Note that the total number of these contacts per N_0 is around unity and hence our secondary bond can be regarded as one of these computed bonds. Although the energetic differences, such as the length and the type of bond, are not at all taken into account, our secondary bond can be considered roughly as an average of computed secondary bonds.

4.2. Bond energy

In table 6 the value of ϵ averaged over all pH values is shown for each protein. Although the pH should affect the thermodynamic parameters to some extent, this is not considered in discussing the difference between proteins. The value of ϵ seems to be around 0.8 kcal/mole, a little higher than RT at room tem-

Table 6
Parameter ϵ and bond energy

	Mgl	Cyr	Chm	Lys	Rna
Value of N_0	269	186	378	215	204
Average value of ϵ	0.787	0.819	0.826	0.884	0.952
Average value of ZJ	1.51	1.57	1.60	1.59	1.56
Number of polar contacts per N_0	0.494	0.376	0.458	0.414	0.397
Number of nonpolar contacts per N_0	0.792	0.731	0.630	0.502	0.441
Number of total contacts per N_0	1.29	1.11	1.09	0.916	0.838
Relative fraction of hydrogen bonds	0.384	0.340	0.421	0.452	0.474
Relative fraction of hydrophobic bonds	0.616	0.660	0.579	0.548	0.526

perature. This energy value must be related to the actual bond energies, hydrogen bond energy and hydrophobic bond energy.

So far a number of authors have estimated the hydrogen bond energy in water. The estimates of the enthalpy change involved in breaking one hydrogen bond and replacing this by the interactions with water range from 0 to 2 kcal/mole [1]. The significant speading of these values may arise from the variety of the environment around the interacting donor and acceptor groups. For example, the presence of nonpolar groups reduces such an enthalpy change, although the total free energy change increases and hence the hydrogen bond gets stronger due to the accompanying formation of hydrophobic bonds [14]. In the absence of hydrophobic bonds, the enthalpy change on formation of a hydrogen bond seems to be around 1.5 kcal/mole.

The van der Waals contribution to the hydrophobic bonds is also incorporated in ϵ of our model. This can be estimated, for example, according to the Lennard–Jones potential function. Némethy and Scheraga [8] used the values of 0.15 kcal/mole for a pair of aliphatic groups and 0.50 kcal/mole for a pair of aromatic groups.

The following procedure gives an estimation of the energies of hydrogen bonds and hydrophobic bonds from the difference in ϵ between five proteins. First let us assume that our values of total bonds N_0 are correct and that the numbers of hydrogen and hydrophobic bonds are expressed by the relative fraction of computed contacts shown in table 6. Then, by the relation

$$\epsilon = \epsilon_H n_H + \epsilon_{H\phi} n_{H\phi} , \qquad (18)$$

where n_H and $n_{H\phi}$ are the relative fraction of hydrogen and hydrophobic bonds, respectively, the energy of a hydrogen bond ϵ_H and that of a hydrophobic bond $\epsilon_{H\phi}$ can be estimated from table 6. The least squares fitted values are: $\epsilon_H = 1.4$ kcal/mole and $\epsilon_{H\phi} = 0.44$ kcal/mole. These results are in good agreement with previous estimations described above. Note that $\epsilon_{H\phi}$ contains only the change in van der Waals interaction and the contribution from the water structure can be estimated as $\Delta g_w / n_{H\phi} = 0.7$ kcal/mole at room temperature, which yields a total free energy of a hydrophobic bond not so different from that of a hydrogen bond. This fact seems to support our "identical and uniform" point of view on the protein structure.

The number of total contacts computed from the tertiary structure increases with increasing hydrophobicity as shown in table 6. If this difference is significant, an alternative procedure must be followed to estimate energy values. As was pointed out above, the value of $N_0\epsilon$ is nearly constant for a small variation of N_0 around the value shown in table 6. Therefore, if we use as N_0 the thus computed number of total contacts, which is larger for myoglobin and smaller for ribonuclease than our actual N_0 values, ϵ values must be proportionally smaller or larger. The least squares fitting of thus converted data yields the values: $\epsilon_H = 3.2$ kcal/mole, $\epsilon_{H\phi} = -0.46$ kcal/mole. However, the computed number of interacting pairs cannot be regarded as the "actual" number of hydrophobic bonds, since an "actual" hydrophobic bond would be formed by the adhesion of more than two nonpolar groups. Nevertheless, a relatively higher contribution of hydrogen bonds than van der Waals interactions is apparent.

4.3. Cooperative energy

The cooperative energy ZJ, which mainly reflects the connectivity of the chain, has been determined as 1.5 to 1.6 kcal/mole and is nearly constant for all the proteins studied (table 6).

4.4. Chain entropy

The chain entropy α of completely disordered state ($X = -1$) is around 4 eu per secondary bond, which is equivalent to 6.6 eu per residue. Previous estimations of the increase in rotational freedom during denaturation are in the range 2 to 6 eu per residue [1]. Suppose that there are some single bonds about which rotation can occur and that they are completely frozen in when the secondary bonds are in the bonded state and are completely released in the unbonded state. The torsional potential function of the C–C or C–N bond in a small molecule has usually 3 minima. In a protein molecule, however, on account of the steric restrictions and the existence of other atomic groups, the number of permitted configurations must be considerably smaller. If we assume 4 such single bonds per residue, the value of 6.6 eu corresponds to

Table 7
Parameter α and chain entropy

	Mgl	Cyr	Chm	Lys	Rna
Average value of α	3.94	3.95	3.85	3.87	4.01
Value of α per residue, $m\alpha$	6.93	7.06	6.04	6.45	6.60
Statistical weight, $e^{m\alpha/R}$	32.7	35.1	20.9	25.7	27.7
Number of internal rotations per residue, w	4.74	4.68	4.37	4.48	4.54
Number of minima of rotational potential, x ($x^w = e^{m\alpha/R}$)	2.09	2.14	2.00	2.06	2.08
Number of S–S bonds	–	–	5	4	4
Number of heme groups	1	1	–	–	–

2.3 minima, since $2.3^4 = e^{6.6/R}$.

The actual number of single bonds about which rotation can occur such as C–C, C–O, C–N, and C–S, is found from the known primary structures for these proteins and is shown in table 7. From this we can estimate the number of minima of the rotational potential, or the number of possible configurations taken by one single bond as above. The result shows almost the same values, a little higher value than 2 for all the proteins considered. It should be noted that the value of α represents the total chain entropy divided by N_0 between the strictly rigid conformation at 0 K and the completely random coil conformation. Hence, if the amino acid chain is cross-linked by S–S bridges, the conformational entropy must be fairly restricted. According to a rough estimation of such loop entropy in the random coil state of 250 residues, the number of potential minima is reduced from 3 to 2.87, 2.77, 2.69, 2.62, and 2.55 for 1, 2, 3, 4, and 5 disulfide bridges, where we have assumed that the chain is Gaussian and that the residues are equally divided by the bridges. However, our results in table 7 do not suggest any significant difference between the proteins with and without S–S bonds. This is because the heme group in myoglobin and cytochrome c would restrict the conformational entropy to the same extent as the S–S bridges.

4.5. Hydrophobicity

The value of γ or the number of water molecules affected by a change in state is from 2 to 4 per secondary bond, which is roughly 5 to 10 per hydrophobic bond if we assume 40% of hydrophobic bonds. Némethy and Scheraga [8] calculated such a number from a molecular model, or as high as requirements of packing will permit, on forming a hydrophobic bond of maximum strength between two isolated side chains and obtained a value ranging from 4 to 12. This coincidence arises from the following two reasons. In actual proteins hydrophobic bonds are often formed by the collection of more than two nonpolar groups,

Table 8
Parameter γ and hydrophobicity

	Mgl	Cyr	Chm	Lys	Rna
Average value of γ	3.76	3.70	3.00	2.89	2.54
Number of nonpolar contacts within 4.0 Å per 100 g	1.19	1.10	0.944	0.755	0.662
Number of nonpolar contacts within 4.5 Å per 100 g	2.55	2.40	2.33	2.11	1.54
Number of nonpolar groups per residue [a]	1.38	1.11	1.26	1.06	1.00
Dependence of T_d on GuHCl concentration (°C/M)	24.0	20.4	16.5	13.3	12.3
Average value of ϵ/α	0.200	0.208	0.214	0.229	0.237
Average value of γ/α	0.956	0.938	0.779	0.748	0.631
Ratio of $C(T_d)/A(T_d)$ at $A(T_d) = 0.5$ in fig. 3	1.12	1.16	1.15	1.18	1.20

a) Calculated from the amino acid composition by the present authors. Number of nonpolar groups are: Ala, 1; Val, 3; Leu, 4; Ile, 4; Met, 4; Cys, 2; Pro, 3; Phe, 7.

Table 9
Midpoints of the transition at neutral pH region

Protein	Thermal	GuHCl (25°C)	Reference
Myoglobin	85°C	2.5 M	Puett [29]
Cytochrome c	78	2.6	Knapp and Pace [30]
α-Chymotrypsin	58	2	Greene and Pace [31]
Lysozyme	78	4	Aune and Tanford [32]
Ribonuclease A	62	3	Salahuddin and Tanford [21]

hence the number of water molecules excluded would be larger than the one calculated for pair interaction of isolated groups. Oppositely, in the denatured state nonpolar groups would not be "fully" in contact with water due to steric restrictions, which would result in the decrease of this number.

The value of γ is better related to the number of nonpolar contacts calculated by Privalov and Khechinashvili [13] than to the number of nonpolar groups themselves, such as CH_3, CH_2 and CH, calculated by us, as shown in table 8.

Let us discuss other expressions for the hydrophobicity than the one derived from structurally determined parameters. In table 9, the midpoint of the transition is tabulated for the thermal denaturation and for GuHCl denaturation at 25°C both in neutral pH region. The latter is defined as the GuHCl concentration which produces the thermal denaturation with $T_d = 25°C$. Then the change in T_d due to the addition of 1 mole of GuHCl is estimated and shown in table 8. The result is quite satisfactory. The more hydrophobic a protein is, the more affected it is in its stability by the presence of denaturants. The efficiency of denaturants can be also expressed by the ratios ϵ/α and γ/α as will be shown later. The smaller the value of ϵ/α and the larger the value of γ/α, the more hydrophobic the protein, which agrees with what is obtained.

As described in the paper I, the ratio C/A expresses the extent of the two-state like nature of the transition. From eqs. (14) and (15) it follows that

$$\frac{C(T_d)}{A(T_d)} = \frac{s(T_d)}{2R}\,\frac{ZJ}{h(T_d)}, \tag{19}$$

and can be considered analogously in the present case. This value is calculated from the parameters as they were determined and shown in fig. 3. The results suggest that all the proteins studied have the value $C/A > 1$ and hence undergo a two-state like structural transition. Moreover, the less hydrophobic proteins have a higher C/A (fig. 3 and table 8), or are of more two-state-like nature. According to the kinetic theory of our model, which will be published later, two rate constants characterize a protein. The faster one is related to the shift in the peak of the population and the slower one to the passing of the population over the potential barrier. A more two-state-like nature means a higher barrier and hence a slower rate. There-

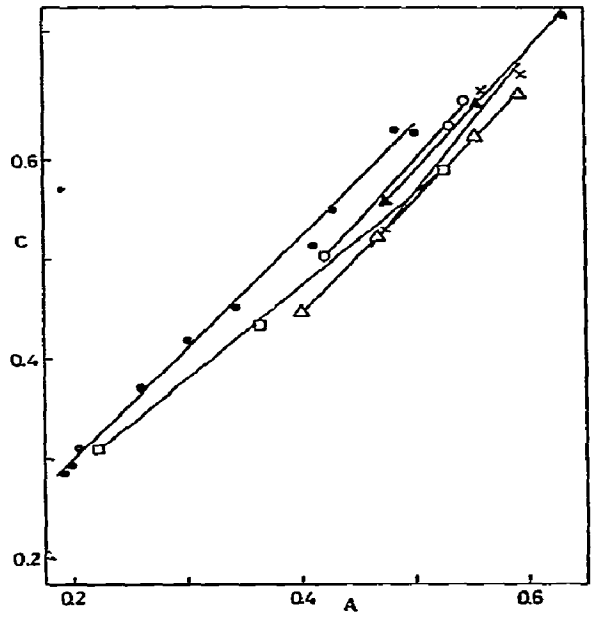

Fig. 3. The two-state likeness can be seen by the position in the parameter space of A and C. A higher value of C/A means a more two-state-like nature. Note that all the proteins in this figure have this ratio larger than 1 and hence undergo the all-or-none-like structural transition. •, ribonuclease A [12]; ○, ribonuclease A; ▲, lysozyme; □, α-chymotrypsin; ×, cytochrome c; △, myoglobin.

fore, the following prediction is deduced; the less hydrophobic a protein is the slower its kinetic constant. Actual kinetic experiments will be analyzed in a forthcoming paper.

5. Effect of pH

In fig. 4, which is plotted from the results of table 2, we can observe the change in the thermodynamic parameters accompanying acidic pH changes for ribonuclease A according to Tsong et al. [12]. All parameters undergo some alteration at around pH 4, where carboxyl groups dissociate. In the low pH region where the net charge is high, the electrostatic repulsion between charged surface groups would reduce the bond energy ϵ and the chain entropy α. The increase in the cooperative energy ZJ would not be significant, since the change is less than 10%.

Let the protein molecule be a rigid, impenetrable sphere of low dielectric constant with the net charge Z uniformly distributed on the surface. The electrostatic energy of this sphere can be calculated (see for example, Tanford [15])

$$\Delta G_{el} = \frac{N_a Z^2 e^2}{2D}\left[\frac{1}{b} - \frac{\kappa}{1+\kappa a}\right], \qquad (20)$$

where N_a is Avogadro's number, D is the dielectric constant of the solvent, κ is the reciprocal of the radius of the ionic atmosphere (Debye length), and b and a are the radius of the sphere and the radius of exclusion respectively. Using the values of $b = 20$ Å, $a = b + 2.5 = 22.5$ Å [15]. $1/\kappa = 9.3$ Å (at ionic strength 0.1), and $D = 80$, and assuming that the value of Z increases from 6 to 17 upon an acidic pH change, the electrostatic contribution is 0.05 kcal/mole per secondary bond. The observed change in ϵ is about 0.3 kcal/mole (fig. 4). However, the use of the bulk dielectric constant of water would not be suitable and the charges are by no means uniformly distributed on the surface in actual proteins. According to Tanford's calculation [16], where the charges were treated as discrete points near the surface of a sphere with dielectric constant 2, the titration curve was quite sensitive to the location of the dissociable sites and he obtained reasonable results only if the charged sites were placed at about 1 Å below the surface of the sphere. Therefore, this estimation should be made considering local electrostatic effects and using an appropriate value of the dielectric constant.

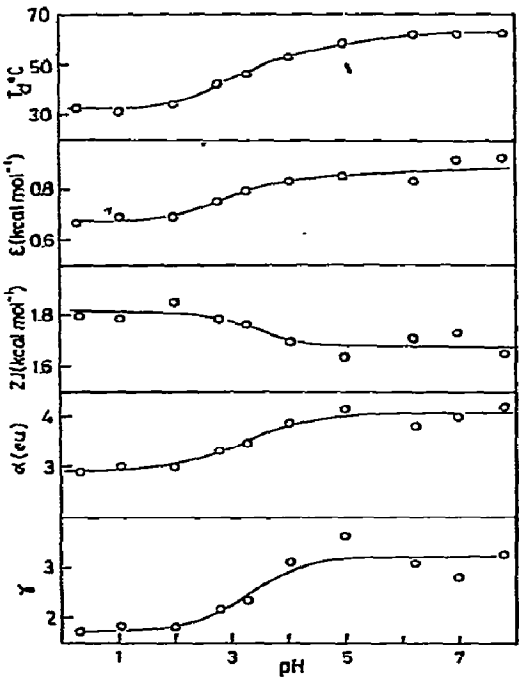

Fig. 4. Thermodynamic parameters determined by our method for ribonuclease A [12] at various values of pH.

The largest decrease with pH is observed in the parameter γ. This must be explained by the change in the water structure around newly discharged groups due to the shift in pH. In fig. 5 the calculated thermal denaturation curves, using the thermodynamic parameters of ribonuclease in the acidic pH region are compared to the actual experimental results by Brandts and Hunt [17]. There is a common behaviour in the post-transition state, that is, the denatured state gets gradually more ordered with the rise of temperature, as well as with the rise of pH at constant temperature. This type of transition is often observed in the presence of denaturants and moreover the curve for pH 1.13 of fig. 5a seems to indicate the tendency to get the parabolic-type transition (fig. 8a), which is sometimes observed at high denaturant concentration. As will be shown in the next section, the effect of the addition of denaturants is well related to the decrease in the free energy change of water Δg_w, or effectively the parameter γ, in eq. (4). Hence, we assume that the

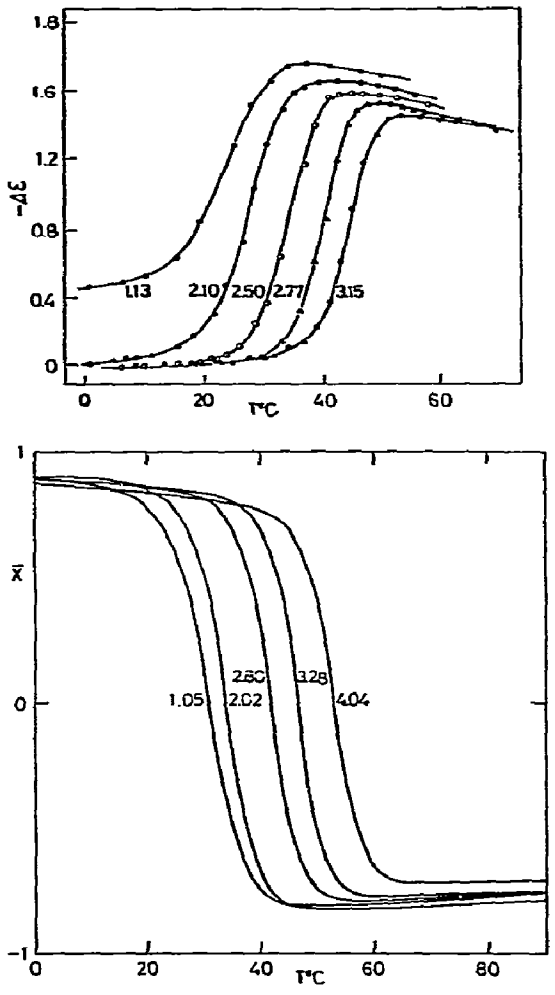

Fig. 5. Effect of pH; (a) the change in the extinction coefficient at 287 nm for ribonuclease A at five pH values [17] (reproduced from Journal of the American Chemical Society). (b) the calculated change of the structural state using the parameters determined from the calorimetric experiment of ribonuclease A [12].

lower pH tends to weaken the hydrophobic interactions due to the change in the water structure around the newly discharged groups. The relation between the parameter γ and the hydrophobic interactions will be discussed later in more detail.

6. Effect of solvent composition

A variation of the solvent composition causes structural changes of the proteins. However, there are no precise calorimetric measurements available in the presence of denaturants or salts. Therefore, we only shall give a semi-quantitative discussion using a number of optical measurements. Since a least squares determination of all the thermodynamic parameters from optical data alone has failed owing to the difficulty that the optimization function (17) has a long and narrow valley in the parameter space of P [18], only the change in the denaturation temperature accompanying the variation of the solvent composition is used for the present analysis.

It is widely known that many protein denaturants, such as urea and GuHCl, owe their action, at least for a large part, to hydrophobic effects. Therefore, we assume that the free energy difference of the water molecule Δg_w is altered in the presence of other substances. In practice we further assume that the temperature dependence of Δg_w is not affected by these. Hence, the denaturants are supposed to reduce effectively the value of γ in eq. (4), without any alteration of the parameters a, b, and c. Fig. 6 shows the temperature dependence of the structural state $\bar{X}$ at various values of γ, which represents the thermal denaturation of a protein at various values of denaturant concentration c. Although the explicit form of $\gamma(c)$ is not clear, it may be a monotonic function of c. Decreasing values of γ in this figure should therefore correspond to increasing values of c. Fig. 6 is replotted in fig. 7 as the isothermal denaturation by denaturant at various values of temperature. The case of $\gamma = 3.2$ in fig. 6 corresponds to $c = 0$ in fig. 7, and $\gamma(c)$ is treated as a linearly decreasing function of c.

Qualitatively, our sample calculation shown in figs. 6 and 7 has the following three aspects similar to what is actually observed in the presence of denaturants by a number of authors, such as Pace and Tanford [19], Tanford and Aune [20], Salahuddin and Tanford [21], and Brandts and Hunt [17], as reproduced in figs. 8a–d. First, the termal denaturation curves at high denaturant concentration of figs. 8b–d are similar to these of fig. 8a, although this is not observed in the low temperature region. Therefore, such a parabolic-type structural change as shown in fig. 8a would be common to all proteins, although it is not always observed at room temperature, and the overall structural changes at various denaturant concentrations would be as those in fig. 6. Secondly, it is often ob-

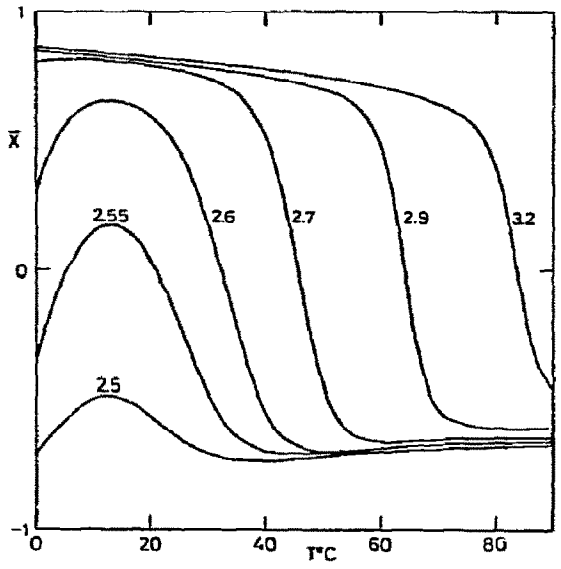

Fig. 6. Thermal denaturation curves with various values of γ calculated from the free energy of eq. (4). Parameters are $N_0 = 250$, $\epsilon = 0.7$ kcal/mole, $ZJ = 1.6$ kcal/mole, $\alpha = 3.35$ eu, $\gamma = 3.2$ to 2.5, $a = -9.35 \times 10^{-1}$ kcal/mole, $b = 5.93 \times 10^{-3}$ kcal/(deg mole), and $c = -8.06 \times 10^{-6}$ kcal/(deg^2 mole).

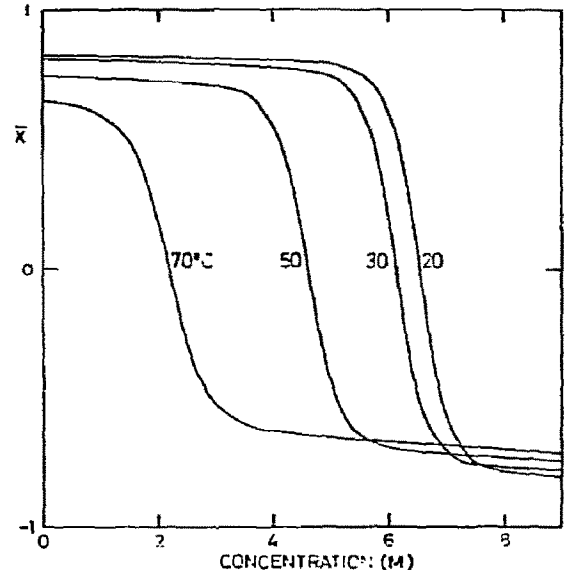

Fig. 7. Isothermal denaturation curves by protein denaturant at various values of temperature. The parameters are the same as those in fig. 6. Denaturant concentration is arbitrarily scaled and is transformed to γ as $\gamma(c) = 3.2 - 0.1\,c$, where c is the concentration of denaturant.

served in the presence of denaturants (fig. 8) that the structure gets gradually more ordered with increasing temperature in the post-transition state, which can be seen also in the sample calculation of fig. 6. Thirdly, the isothermal denaturation by denaturants usually yields a more disordered state than the thermally induced one [3]. This would be explained by the fact that usual proteins have values of A and C around 0.5 (fig. 3) and hence the remaining order in the post-transition state, which is specified by the position of the characteristic point P($-C$, $-A$) in fig. 1 of the paper I, is fairly large. On the other hand the addition of denaturants effectively shifts the characteristic point toward the region K in that figure, and hence causes a more disordered structure.

The change in the denaturation temperature T_d accompanying the addition of denaturants can be converted to the change in the parameter γ by the following procedure. Provided that the parameters ϵ and α are not altered in the presence of denaturants, the value of γ is obtained from the first derivative of eq. (4),

$$\gamma = -\frac{\epsilon - \alpha T_d}{a + bT_d + cT_d^2}. \tag{21}$$

Experimental observations of the change in the denaturation temperature for some proteins with some denaturants may be converted to the changes in parameter γ according to eq. (21). Numerical values of a, b, and c are listed in table 4 and the value of the parameters ϵ and α are either taken as the average of those in table 3, or else assumed reasonably with the value of ϵ/α estimated from the "hydrophobicity temperature" (see below). The result is normalized by the value at zero denaturant concentration and is shown in fig. 9.

It can be seen that for several proteins a single slope is obtained in a given denaturant. As is well known GuHCl is a stronger denaturant than urea, hence the slope is steeper. Since the change in γ represents essentially the change in Δg_w, or the free energy change of the water structure, in our point of view this can be connected with the experimentally observable free energy of transfer from water to aqueous solutions

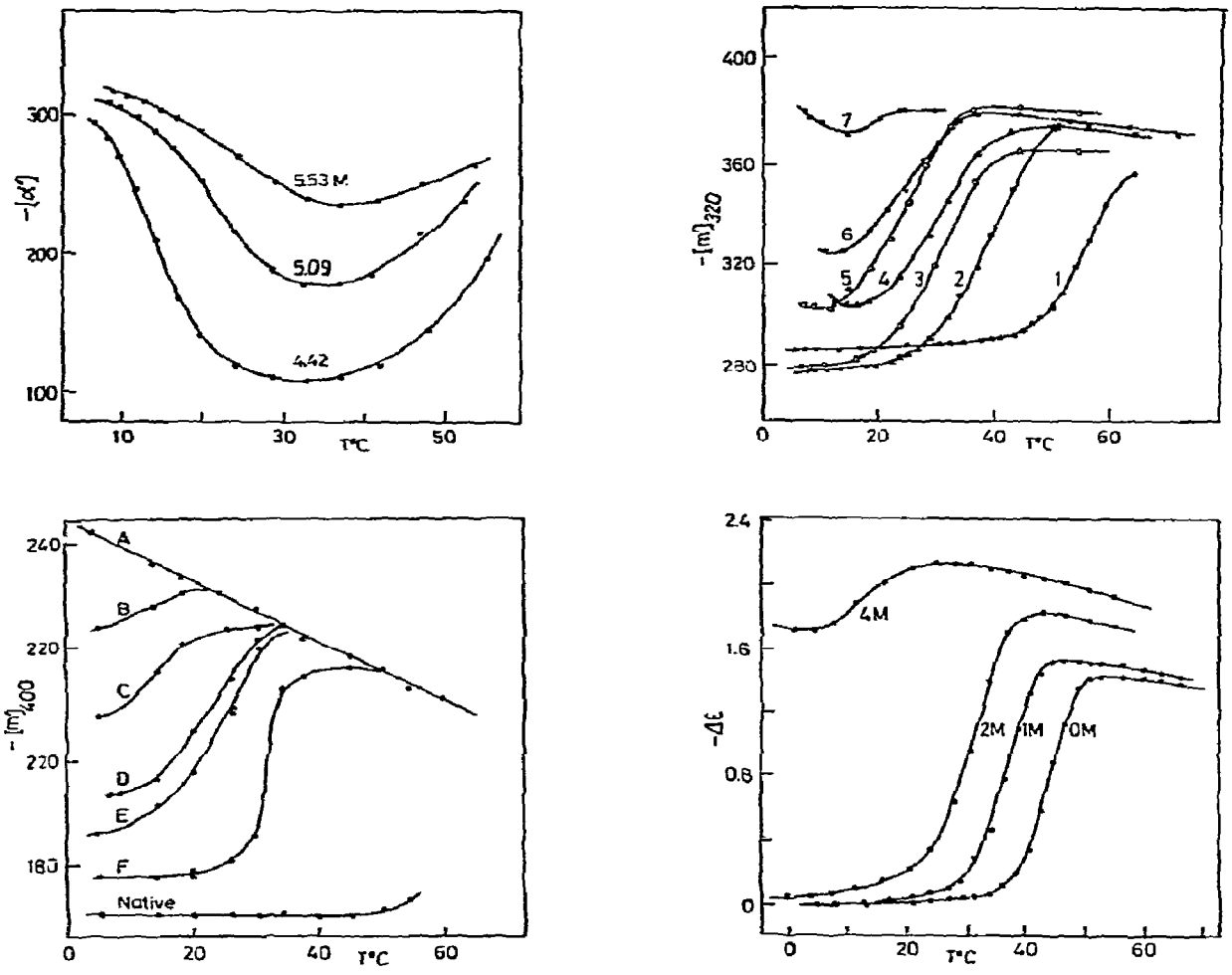

Fig. 8. Effect of solvent composition; (a) the optical rotation at 365 nm for β-lactoglobulin at various urea concentration and at pH 2.5 [19] (reproduced from Biochemistry); (b) the optical rotation at 320 nm for lysozyme at various GuHCL concentration and at pH 2 to 3 [20]; 1–7 denote 0.85, 1.92, 2.23, 2.79, 3.08, 3.14, and 3.50 M, respectively; (reproduced from Biochemistry); (c) the optical rotation at 400 nm for ribonuclease at various GuHCL concentration and at pH 6 [21]; A–F denote 4.31, 3.44, 3.24, 3.10, 2.79, and 2.29 M, respectively; (reproduced from Biochemistry); (d) the change in the extinction coefficient at 365 nm for ribonuclease at various urea concentration and at pH 3 [17] (reproduced from Journal of the American Chemical Society).

containing denaturants. A large amount of such data has been obtained by Tanford's group [3] and they are in agreement with our observation. These facts seem to support our assumption that the effect of denaturants is mainly explained by the parameter γ.

It is well-known that in pH region away from the isoelectric point, where repulsive electrostatic interactions predominate, the effect of salt will normally tend to reduce these repulsions and hence stabilize the more compact conformation. However, at high concentration the effect of salt must be interpreted by the short-range interactions between salt ions and water structure rather than the electrostatic one. Thus, the thermal denaturation of some proteins with NaCl is analyzed in the same manner as above and shown in the same figure. The slope is positive for this salt which stabilizes the protein. The effect is satisfactory and again seems to support our point of view for the

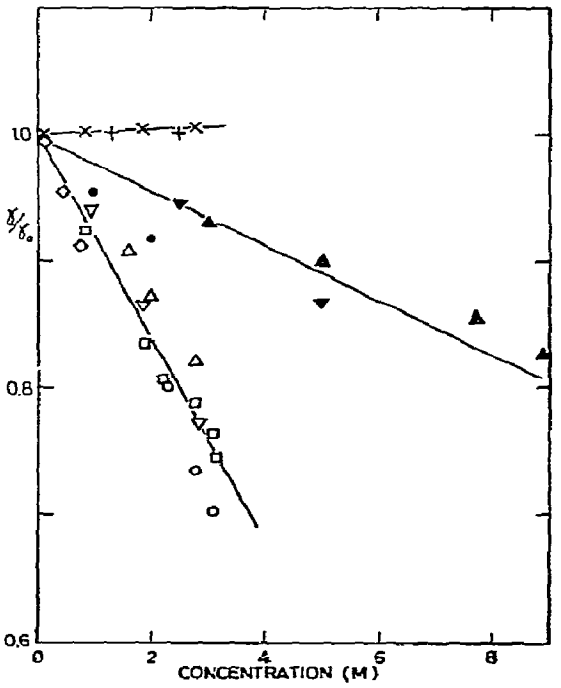

Fig. 9. Calculated free energy of transfer or the relative change in parameter γ due to the addition of denaturant or salt. ○, ribonuclease in GuHCL [21]; ▽, ribonuclease in GuHCL [23]; ●, ribonuclease in urea [17]; ▼, ribonuclease in urea [23]; +, ribonuclease in NaCl [23]; □, lysozyme in GuHCl [20]; △, lysozyme in GuHCl [24]; ▲, lysozyme in urea [24]; ◇, flagellin in GuHCl [25]; ×, flagellin in NaCl [25].

effect of solvent composition even for salt.

Before terminating the discussion of this section, let us point out a practical measure of the hydrophobicity obtainable from experiment. It is the temperature when the parabolic-type relation between $\bar{X}$ and T is observed by the addition of a denaturant. Note that sometimes there exist two denaturation temperatures or no denaturation temperature at all in the case of such a parabolic-type transition of fig. 8a. As an index for the existence of this type of transition at room temperature, the condition $(\partial\gamma/\partial T_d)_{T_d=T_d^h} = 0$, or the relation

$$\epsilon(b + 2cT_d^h) + \alpha(a - c(T_d^h)^2) = 0\,, \qquad (22)$$

is used, where the denaturation curve is tangent to the axis of X. In our sample calculation of fig. 6, this "hydrophobicity temperature" is $T_d^h = 12.8°C$, which corresponds to $\gamma^h = 2.53$.

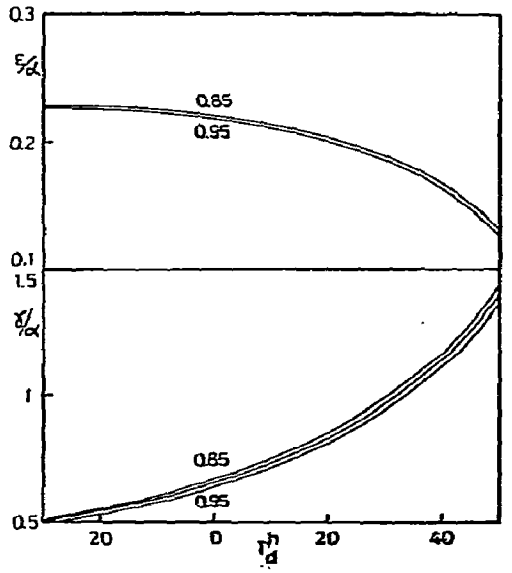

Fig. 10. The ratios ϵ/α (kcal mole^{-1}eu^{-1}) and γ/α (eu^{-1}), versus the hydrophobicity temperature T_d^h, calculated according to eqs. (21) and (22) with the values of a, b, and c, for plausible relative aliphatic contents, $w_Q/(w_Q+w_T) = 0.85, 0.9$, and 0.95 (see table 4).

The hydrophobicity temperature characterizes the hydrophobic nature of a protein. According to eqs. (21) and (22), two ratios, ϵ/α and γ/α, are plotted against T_d^h in fig. 10 for the parameters a, b, and c determined from plausible values of the relative aliphatic content (see table 4). It can be seen that the parameter γ, or the hydrophobicity gets larger as the T_d^h gets higher. Therefore, lysozyme seems to be more hydrophobic than ribonuclease and β-lactoglobulin is the most hydrophobic as can be inferred from fig. 8. Although the difference in pH cannot be ignored, the result is in the same direction as that obtained in sect. 4 by calorimetric data and by the calculation of the extent of hydrophobicity (table 8).

7. Discussion

Thus far we have discussed several aspects of the structural stability of globular proteins within the framework of our simplified model. The structure of a globular protein has been considered to be main-

tained by identical secondary bonds uniformly distributed and by interactions between the chain and water. This "identical and uniform" assumption of the model is extensively supported by the fact that the average values of the thermodynamic parameters ϵ, α, and γ determined are quantitatively explained by the molecular properties and the tertiary structures of individual proteins. Estimated energy values of hydrogen bonding and van der Waals interactions correspond well to reported values. The average number of rotational minima around a single bond is nearly the same for all the proteins studied and is reasonably understood by the steric restrictions. The cooperative energy ZJ which reflects mainly the connectivity of the chain also has the same value for each protein.

Although the thermodynamic parameters ϵ, α, and γ have been revealed to be related to several aspects of the stability of globular proteins, the most significant parameter seems to be γ. Two different physical properties or concepts are included in this parameter. One is the original concept described in the basic assumption of the model, that is, the average number of water molecules which establish new contacts with the nonpolar groups of a protein as a consequence of the breakage of one secondary bond. Therefore, it is directly related to the tertiary structure of a protein. Actually the value of γ is well related to the number of contacts between nonpolar groups rather than the number of nonpolar groups itself (table 8). It may be that through this parameter each protein shows its individuality in the thermodynamic stability and possibly in the kinetic function under our "identical and uniform" assumption.

In addition to the original concept, the change in the free energy difference of the water structure Δg_w due to the change in the solvent composition is conveniently expressed by the change of the parameter γ. Many factors are expected to affect the water structure and the protein stability. The addition of denaturants, the addition of salts, and even the change of pH are considered to be such factors and explained in a unified manner. Similarly we expect that the effect of high pressure on the structure of a protein can be explained mainly through the change of water structures in such an environment.

Although we have used an oversimplified model, the results of our analysis yields valuable information and ideas on the stability of proteins and our model seems to be useful for further investigations. To our regret, however, not enough experimental data are available at present. More precise calorimetric measurements for various proteins and under various external conditions are needed and may then hopefully be analyzed by our model.

Appendix. Vibrational contribution to the absolute specific heat of proteins

It is well known that the infrared and Raman spectra of a certain atomic group are not so dependent on its environment. We therefore have attempted to assign certain wave numbers of vibration to all the internal degrees of freedom of 20 amino acids constituting a protein using the infrared and Raman spectra for small molecules [22].

The internal degrees of freedom for 20 amino acids are itemized in table 10 and the wave numbers assigned are listed in table 11. Some unknown degrees of freedom (column 5 of table 10) are interpreted as those of backbone bending, of which the wave number is assumed to be 450 cm^{-1}. The internal rotations (torsional oscillations) may be frozen-in almost completely in a native form by the formation of secondary bonds and by the steric effects, but become free on denaturation. Therefore, corresponding degrees of backbone bending may change to those of internal rotations (column 6 of table 10), the wave number of which is assumed to be 100 cm^{-1}.

The specific heat for each vibration is given by the well-known relation

$$C_{vib} = \left(\frac{h\nu}{kT}\right)^2 \frac{e^{h\nu/kT}}{(e^{h\nu/kT}-1)^2}. \qquad (23)$$

The absolute specific heats for five globular proteins with known amino acid compositions are thus calculated and shown in table 12.

Table 10
Itemization of the internal degrees of freedom in amino acids [a]

						Bond length stretching											Bond angle bending								
a.a.	n	3n	f	3n−f	r	C–C	C–N	C–O	C–S	S–S	C=C	C=N	C=O	C–H	N–H	O–H	S–H	CH_3 (5)	CH_2 (4)	CH (2)	NH_2 (4)	NH (2)	COOH (2)	OH (2)	SH (2)	RING (12)
Main	5	15	9	6	2	1	1								1											
Gly	2	6	6	0	0									2					1							
Ala	5	15	12	3	1	1								4				1		1						
Val	11	33	25	8	3	3								8				2		2						
Leu	14	42	32	10	4	4								10				2	1	2						
Ile	14	42	32	10	4	4								10				2	1	2						
Phe	15	45	32	13	2	5					3			8					1	1						5/6
Tyr	16	48	33	15	3	5		1			3			7		1			1	1				1		4/6
Trp	19	57	39	18	2	6	2				4			8	1				1	1						6/6
Pro	10	30	25	5	0	3	1							7					3	1						
His^+	11	33	22	11	2	2	3				1	1		5					1	1						2/6
Ser	6	18	14	4	2	1		1						3		1			1	1				1		
Thr	9	27	20	7	3	2		1						5		1		1		2				1		
Met	12	36	27	9	3	2			2					8				1	2	1						
CysH	6	18	14	4	2	1			1					3			1		1	1					1	
$\frac{1}{2}$Cys	5	15	11.5	3.5	2	1			1	1/2				3					1	1						
Asn	9	27	19	8	2	2	1						1	3	2				1	1	1					
Gln	12	36	26	10	3	3	1						1	5	2				2	1	1					
Asp	8	24	16	8	2	2		1					1	3		1			2	1			1			
Glu	11	33	23	10	3	3		1					1	5		1			2	1			1			
Lys	16	48	38	10	5	4	1							9	2				4	1	1					
Arg	18	54	39	15	5	3	3					1		7	4				3	1	1	2				

[a] Main denotes main chain and other amino acids include only side chains. Numbers in parentheses for bond angle bending are the total degrees of freedom for individual atomic groups.
Abbreviations: a.a., amino acids; n, number of atoms; f, degrees of freedom assigned; r, number of single bonds around which rotation can occur.

Table 11
Wave number (cm^{-1}) for each vibration used in our calculation

Stretching		Bending	
C–C	1000	Peptide group	
C–N	1000	Amide I	1650
C–O	1000	Amide II	1550
C–S	700	Amide III	1250
S–S	445	Amide IV	630
C=C	1600	Amide V	700
C=N	1650	Amide VI	600
C=O	1700		
C–H	3000		
N–H	3300		
O–H	3300		
S–H	3300		
CH_3 group		NH group	
symmetric bending	1350	bending (×2)	1900
degenerated bending (×2)	1450	COOH group	
rocking (×2)	1000	OH in-plane bending	1300
		OH out-of-plane bending	930
CH_2 group		Alcohol-OH group	
scissor	1430	OH + CH bending	1400
rocking	900		1200
wagging	1250		
twisting	1150		
		SH group	
CH group		bending (×2)	1000
bending (×2)	1250		
		Benzene ring	
NH_2 group		CH in-plane	
scissor	1500	bending (×6)	1200
rocking	670	CH out-of-plane	
wagging	1000	bending (×2)	400
twisting	1000	(×2)	680
		(×2)	850

Table 12
Calculated specific heat per gram of protein (cal g^{-1} deg^{-1})

Protein	Native		Denatured	
	300 K	350 K	300 K	350 K
Cytochrome c	0.252	0.304	0.277	0.323
Ribonuclease	0.246	0.295	0.270	0.314
Lysozyme	0.249	0.299	0.273	0.317
Myoglobin	0.254	0.306	0.279	0.325
α-Chymotrypsin	0.250	0.301	0.274	0.320
Average	0.250	0.301	0.275	0.320

References

[1] J.F. Brandts, in: Structure and Stability of Biological Macromolecules, ed. S.N. Timasheff and G.D. Fasman (Marcel Dekker, Inc., New York, 1969) p. 213.
[2] R. Lumry and R. Biltonen, in: Structure and Stability of Biological Macromolecules, ed. S.N. Timasheff and G.D. Fasman (Marcel Dekker, Inc., New York, 1968) p. 65.
[3] C. Tanford, Advan. Protein Chem. 24 (1970) 1.
[4] W. Kauzmann, Advan. Protein Chem. 14 (1959) 1.
[5] H.S. Frank and M.W. Evans, J. Chem. Phys. 13 (1945) 507.
[6] M.F. Perutz, European J. Biochem. 8 (1969) 455.
[7] I.M. Klotz, Science 128 (1958) 815.
[8] G. Némethy and H.A. Scheraga, J. Phys. Chem. 66 (1962) 1773.
[9] G. Némethy and H.A. Scheraga, J. Chem. Phys. 15 (1962) 3382.
[10] G. Némethy and H.A. Scheraga, J. Chem. Phys. 15 (1962) 3401.
[11] R. Fletcher and M.J.D. Powell, Computer J. 6 (1963) 163.
[12] T.Y. Tsong, R.P. Hearn, D.P. Wrathall and J.M. Sturtevant, Biochem. 9 (1970) 2666.
[13] P.L. Privalov and N.N. Khechinashvili, J. Mol. Biol. 86 (1974) 665.
[14] H.A. Scheraga, in: The Proteins, Vol. 1, ed. H. Neurath (Academic Press, New York, 1963) p. 477.
[15] C. Tanford, J. Amer. Chem. Soc. 72 (1950) 441.
[16] C. Tanford, J. Amer. Chem. Soc. 79 (1957) 5340.
[17] J.F. Brandts and L. Hunt, J. Amer. Chem. Soc. 89 (1967) 4826.
[18] M.I. Kanehisa, Ph.D. Dissertation, Univ. of Tokyo (1974).
[19] N.C. Pace and C. Tanford, Biochem. 7 (1968) 198.
[20] C. Tanford and K.C. Aune, Biochem. 9 (1970) 206.
[21] A. Salahuddin and C. Tanford, Biochem. 9 (1970) 1342.
[22] S. Mizushima and T. Shimanouchi, Infrared Absorption and Raman Effect (Kyoritsu Shuppan, Tokyo, 1958) [in Japanese].
[23] P.H. von Hippel and K.Y. Wong, J. Biol. Chem. 240 (1965) 3909.
[24] K. Hamaguchi and H. Sakai, J. Biochem. (Tokyo) 57 (1965) 721.
[25] Y. Takayama, M.S. Dissertation, Univ. of Tokyo (1973).
[26] N. Greenfield and G.D. Fasman, Biochem. 10 (1969) 4108.
[27] J.T. Edsall, in: The Encyclopedia of Biochemistry, ed. R.J. Williams and E.M. Lanford, Jr., (Reinhold Pub. Corp., New York, 1967) p. 682.
[28] I.D. Kuntz and W. Kauzmann, Advan. Protein Chem. (1974) 239.
[29] D. Puett, J. Biol. Chem. 248 (1973) 4623.
[30] J.A. Knapp and C.N. Pace, Biochem. 13 (1974) 1289.
[31] R.F. Green, Jr. and C.N. Pace, J. Biol. Chem. 249 (1974) 5388.
[32] K.C. Aune and C. Tanford, Biochem. 8 (1969) 4579.