PROTEINS: Structure, Function, and Genetics 19:291–301 (1994)

Molecular Basis of Cooperativity in Protein Folding. V. Thermodynamic and Structural Conditions for the Stabilization of Compact Denatured States

Dong Xie and Ernesto Freire

Departments of Biology and Biophysics and the Biocalorimetry Center, The Johns Hopkins University, Baltimore, Maryland 21218

ABSTRACT **The heat-denatured state of proteins has been usually assumed to be a fully hydrated random coil. It is now evident that under certain solvent conditions or after chemical or genetic modifications, the protein molecule may exhibit a hydrophobic core and residual secondary structure after thermal denaturation. This state of the protein has been called the "compact denatured" or "molten globule" state. Recently is has been shown that α-lactalbumin at pH < 5 denatures into a molten globule state upon increasing the temperature (Griko, Y., Freire, E., Privalov, P.L. Biochemistry 33:1889–1899, 1994). This state has a lower heat capacity and a higher enthalpy at low temperatures than the unfolded state. At those temperatures the stabilization of the molten globule state is of an entropic origin since the enthalpy contributes unfavorably to the Gibbs free energy. Since the molten globule is more structured than the unfolded state and, therefore, is expected to have a lower configurational entropy, the net entropic gain must originate primarily from solvent related entropy arising from the hydrophobic effect, and to a lesser extent from protonation or electrostatic effects. In this work, we have examined a large ensemble of partly folded states derived from the native structure of α-lactalbumin in order to identify those states that satisfy the energetic criteria of the molten globule. It was found that only few states satisfied the experimental constraints and that, furthermore, those states were part of the same structural family. In particular, the regions corresponding to the A, B, and C helices were found to be folded, while the β sheet and the D helix were found to be unfolded. At temperatures below 45°C the states exhibiting those structural characteristics are enthalpically higher than the unfolded state in agreement with the experimental data. Interestingly, those states have a heat capacity close to that observed for the acid pH compact denatured state of α-lactalbumin [980 cal (mol·K)$^{-1}$]. In addition, the folded regions of these states include those residues found to be highly protected by NMR hydrogen exchange experiments. This work represents an initial attempt to model the structural origin of the thermodynamic properties of partly folded states. The results suggest a number of structural features that are consistent with experimental data.**

Key words: protein folding, cooperativity, folding intermediates, protein thermodynamics

INTRODUCTION

It is now evident that under certain solvent conditions, the heat-denatured state of some proteins does not resemble a fully hydrated random coil but rather a compact state exhibiting a significant hydrophobic core and different degrees of residual secondary structure (see refs. 1–3 for recent reviews). This state has been often referred to as a "compact denatured" or "molten globule" state. In this paper, both terms will be used interchangeably and will denote a protein state characterized by (1) significant secondary structure, (2) significant compactness due to the presence of a sizable hydrophobic core, and (3) a tertiary structure reminiscent of the native fold but without necessarily exhibiting the packing of the native state. These characteristics have been discussed elsewhere.[4]

Recently, it has been shown for the case of α-lactalbumin that at low pH (<5) this protein denatures into a molten globule state and that this state progressively unfolds upon increasing the temperature.[5] At progressively lower pH values the denaturation transition occurs at increasingly lower temperatures until, at pH values lower than 3, the transition is no longer visible and the molten globule state becomes the predominant species. The state obtained by temperature denaturation is spec-

Received November 29, 1993; revision accepted May 2, 1994.
Address reprint requests to Ernesto Freire, Departments of Biology and Biophysics and the Biocalorimetry Center, The Johns Hopkins University, Baltimore, MD 21218.

troscopically identical to the acid denatured state or "A" state of α-lactalbumin. At 100°C the heat capacity and optical properties of the protein become experimentally indistinguishable from that of the completely unfolded state. Consistent with the presence of a sizable hydrophobic core, the molten globule state has a lower heat capacity than the unfolded state. Also, at low temperatures the molten globule state has a higher enthalpy than the unfolded state. According to the LeChatelier principle, under those conditions a temperature increase cannot induce a transition from the molten globule state to the unfolded state. This situation does not remain the same at all temperatures since the unfolded state has a higher heat capacity than the molten globule state. As the temperature increases, the enthalpy of the unfolded state approaches and eventually surpasses the enthalpy of the molten globule state. The unfolded state becomes the highest enthalpy state and at that point it is stabilized by a temperature increase.

It is becoming increasingly clear that the molten globule state of different proteins is characterized by the persistence of secondary structure elements that preserve the tertiary fold of the native state and, to a significant extent, the native packing.[6–9] In the case of α-lactalbumin in particular, the NMR data suggest that the persistent structure in the acid molten globule is native-like.[7] In general, the exact amount of residual structure is not the same for all molten globules, some of them being almost native-like and other ones closer to the unfolded state.[8] This structural diversity emphasizes the need for an identification of the thermodynamic determinants of this state.

The motivation for this paper was to examine a large ensemble of partly folded states of α-lactalbumin in order to identify those states that satisfy the experimental enthalpy and heat capacity of the compact denatured or molten globule state. Previously, it has been shown that the enthalpy and heat capacity changes between protein states can be predicted accurately using an empirically derived structural parameterization.[4,10] On the other hand, an arbitrarily large number of partly folded states can be generated from the crystallographic or NMR-derived structures by a combinatorial unfolding algorithm. Using this computational procedure in conjunction with the experimental thermodynamic data, it has been possible to generate a family of structural states with thermodynamic properties consistent with those observed experimentally.

STRUCTURAL PARAMETERIZATION OF THE ENTHALPY AND HEAT CAPACITY CHANGES

Previously, we and others have shown that the heat capacity difference between arbitrary conformational states of a protein can be expressed as a linear combination of the differences in polar (ΔASA_{pol}) and apolar (ΔASA_{ap}) solvent accessible surface areas between those states[10–15]:

$$\Delta C_p = 0.45 \cdot \Delta ASA_{ap} - 0.26 \cdot \Delta ASA_{pol} \quad (1)$$

where 0.45 and −0.26 are the elementary contributions per Å^2 of apolar and polar area exposed to water in cal (K·mol)$^{-1}$. All our accessible surface area calculations of protein structures and partly folded states were analyzed as described before[10] using the implementation of the Lee and Richard's algorithm[16] in the program ACCESS (Scott R. Presnell, University of California, San Francisco), with a probe radius of 1.4 Å and a slice width of 0.25 Å. The rest of the analysis was performed with software developed in this laboratory (this software has been written for a Silicon Graphics platform and is available from this laboratory upon request). In Eq. (1), the numerical coefficients have been obtained from a statistical analysis of the proteins for which high resolution thermodynamic and structural information is available, and the thermodynamics of aqueous dissolution of solid dipeptides.[11] It has been shown that Eq. (1) predicts the heat capacity change for unfolding with an accuracy close to the experimental error in the determination of ΔC_p (±9%). In this paper, all thermodynamic quantities will be referenced to the native state unless otherwise noted.

The bulk of the enthalpy change can also be expressed as a linear combination of the changes in polar and apolar solvent accessible surface areas:

$$\Delta H(T) = a(T) \cdot \Delta ASA_{ap} + b(T) \cdot \Delta ASA_{pol} \quad (2)$$

$$\Delta H(T) = a(T_R) \cdot \Delta ASA_{ap} + b(T_R) \cdot \Delta ASA_{pol} + \Delta C_p \cdot (T - T_R) \quad (2a)$$

where T_R is an appropriately chosen reference temperature. In previous papers, we used 100°C as the reference temperature because this temperature is equal to the so-called convergence temperature, at which the residue-normalized enthalpy for most globular proteins assumes a constant value. Recently, however, we have found that a better reference temperature is 60°C, which corresponds to the median transition temperature for the proteins in the database. This choice minimizes extrapolation errors due to uncertainties in ΔC_p and the possibility that ΔC_p for some proteins might not be constant and that it decreases at higher temperatures as suggested by Privalov and Makhatadze.[17] At 60°C the enthalpy change can be written as

$$\Delta H(60) = 31.4 \cdot \Delta ASA_{pol} - 8.44 \cdot \Delta ASA_{ap} \quad (3)$$

Figure 1 illustrates the accuracy with which the enthalpy and heat capacity changes for protein unfolding are accounted for by Eqs. (1) and (3). It is clear that this parameterization provides an accurate estimation of enthalpy values from structural parameters. At 60°C, the average error between the exper-

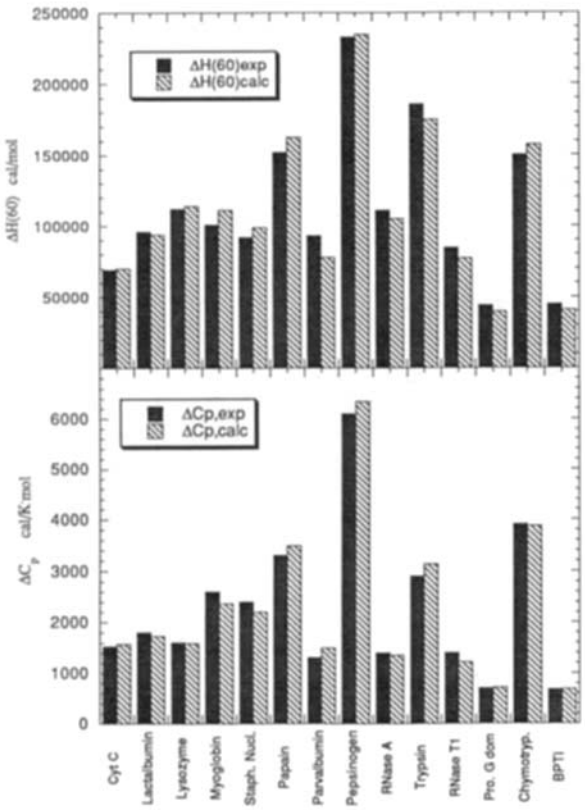

Fig. 1. Comparison between the experimental and calculated heat capacity, ΔC_p, and enthalpy, $\Delta H(60)$, changes for the unfolding of the proteins in the thermodynamic database (see ref. 10 for details).

imental and calculated values is 6%. Under most conditions, Eqs. (2) and (3) account for over 90% of the enthalpy change of unfolding. The additional terms correspond mainly to the enthalpies associated with protonation or the effects of specific ligands, if present. Those additional contributions need to be taken into account explicitly, especially at low temperatures in which the contribution given by Eq. (2) is close to zero. The protonation of carboxylic groups has an enthalpy close to -1 kcal mol^{-1} and that of histidyl groups is close to -7 kcal mol^{-1}. In the case of α-lactalbumin, the enthalpy associated with the binding of Ca^{2+} appears to be small and not to contribute significantly to the enthalpy change of denaturation.[5]

COMBINATORIAL GENERATION OF PARTLY FOLDED STATES

The most significant feature of the experimental energetics of the molten globule state of α-lactalbumin is that it is enthalpically higher than the unfolded state at temperatures lower than 45°C. Since Eqs. (1)–(3) can be used to accurately predict enthalpy and heat capacity changes from structure, it was assumed that the same functions could be used to estimate the corresponding quantities for partly folded states. Therefore, a strategy was devised to generate partly folded states of α-lactalbumin, perform the thermodynamic calculations, and identify those states that satisfy the experimental constraints of the molten globule. For this purpose, partly folded states of α-lactalbumin were generated using the crystallographic structure[18] as a template. This can be accomplished by subdividing the protein into an arbitrary number, N, of folding units, and then generating partly folded states by folding and unfolding the basic folding units in all possible combinations. In this way, a total of 2^N states are created, corresponding to all possible ways of folding and unfolding the N folding units. For these calculations, it is assumed that the unfolded regions are exposed to the solvent and hydrated while the folded regions maintain the conformation existing in the native state. The actual exposure to the solvent and the degree of hydration of folded regions depend on the state of the remaining units in the protein.

In principle, the partitioning of the protein into different folding units can be made using arbitrarily small units, down to the residue level. This partitioning level is, however, computationally intractable since even small proteins will generate over 2^{100} states. A reasonable alternative is to begin the calculations with an *initial* partitioning based on secondary structure elements, since a considerable body of experimental evidence suggests that these elements tend to be either folded or unfolded in partly folded states of proteins (see, for example, refs. 19–21). The resolution of the partitioning grid is then increased by further subdividing the protein into smaller units comprising only parts of secondary structure elements. This process can be repeated to the desired level of resolution. It must be emphasized, however, that this approach does not imply or assume that secondary structure elements are intrinsically stable or that they constitute the earliest folding elements. Within this context, the conceptual differences between stability and cooperativity must be recalled. Cooperativity refers to the absence or presence of equilibrium folding intermediates independently of stability. For example, an isolated, solvent exposed α-helix is usually highly unstable even though its folding/unfolding is highly cooperative (see ref. 22 for a general reference).

In order to test the self-consistency of the above partitioning strategy and confirm that the results obtained this way truly reflect a protein property and not an artifact of the partitioning scheme, different partition alternatives were considered. For this purpose, the number as well as location of the basic folding units were varied. For α-lactalbumin the following major schemes were considered and analyzed:

A: {1–18}, {19–33}, {34–67}, {68–82}, {83–104}, {105–113}, {114–122}
B: {1–18}, {19–37}, {38–46}, {47–67}, {68–82}, {83–104}, {105–113}, {114–122}
C: {1,18}, {19–29}, {30–37}, {38–46}, {47–67}, {68–82}, {83–92}, {93–104}, {105–113}, {114–122}
D: {1–11}, {12–18}, {19–29}, {30–37}, {38–46}, {47–67}, {68–82}, {83–92}, {93–104}, {105–113}, {114–122}
E: {1–11}, {12–18}, {19–29}, {30–37}, {38–46}, {47–67}, {68–82}, {83–92}, {93–99}, {100–104}, {105–113}, {114–122}

in which the numbers in brackets specify the sequence boundaries of each folding unit.

The above partitioning schemes generate from 128 states (scheme A) to 4096 states (scheme E). In addition, upward and downard shifts in the boundaries of the folding units by up to three residues were performed. This was done in order to minimize the effects of uncertainties arising from the definition of the folding unit boundaries. For example, Ser-34 is strictly the last amino acid in helix B, however it is strongly hydrogen bonded to Asp-37, so its location in folding unit 2 or 3 in scheme A is not clear. Also, Asp-82 is the last amino acid in a 3_{10}-helix and also part of the Ca^{2+} binding loop.

Scheme A is the most conservative and uses the most strict sequence definition of the secondary structure elements. In this scheme, the first unit includes the A α-helix (residues 5–11) and an immediately adjacent 3_{10}-helix (residues 12–16). The second unit includes a type III′ turn and the B α-helix (residues 23–34). The third unit consists of the only β-sheet structure which has two strands (residues 40–50) and six type I and III turns (residues 34–39 and 51–67). The fourth unit includes a 3_{10}-helix (residues 76–82) and the two turns before the helix. The fifth unit includes the calcium binding loop, the C α-helix (residues 86–99) and a 3_{10}-helix (residues 101–104). The sixth unit corresponds to the short D α-helix (residues 105–109) and an irregular turn (residues 110–113). Finally, the seventh unit corresponds to a 3_{10}-helix (residues 115–119) and the carboxyl terminal. This partitioning is illustrated in Figure 2. Schemes D and E, on the other hand, do not follow explicit rules and split helices A, B, and/or C in half.

The results obtained with the different partitioning schemes were very similar to each other, i.e., the same regions of the protein were predicted to be folded or unfolded in those states that exhibited enthalpies and heat capacities that were in agreement with the experiment. The main difference was observed for helix A (residues 5–11) and the immediately adjacent 3_{10}-helix (residues 12–16). In this case, it was found that the apolar area buried by the 3_{10}-helix was larger than the corresponding area for

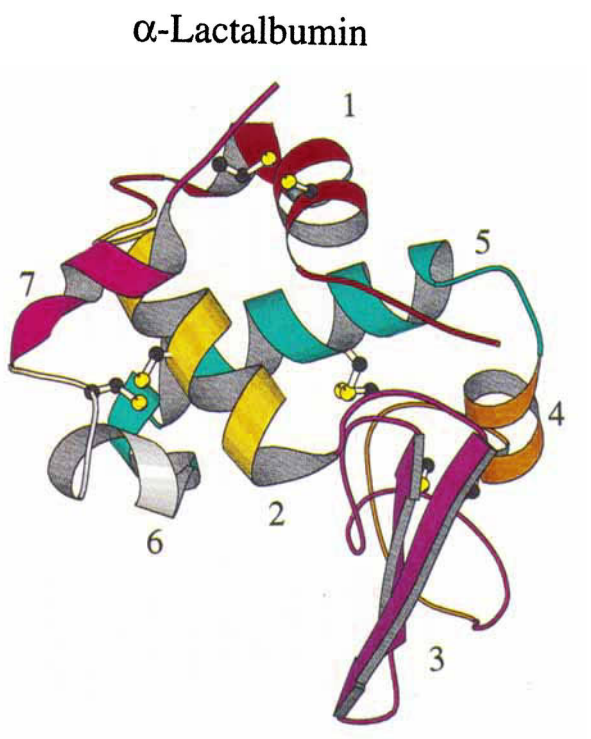

Fig. 2. The partitioning of α-lactalbumin into seven folding units corresponding to unit 1 {1–18} (red), unit 2 {19–33} (yellow), unit 3 {34–67} (purple), unit 4 {68–82} (orange), unit 5 {83–104} (green), unit 6 {105–113} (white), unit 7 {114–112} (magenta). This figure was generated with the program Molscript.[36]

helix A, suggesting that the 12–16 3_{10}-helix might be stabilized by stronger hydrophobic contacts. On the basis of these results and for the clarity and simplicity of presentation, our discussion will be based on the seven folding units partition described above (scheme A). When necessary, mention will be made to the results obtained with higher resolution partitionings.

In all of our calculations, the folded regions in partly folded states are assumed to preserve the conformation existing in the native state. This is certainly an oversimplification; however, it is not expected to introduce a significant error in the calculated heat capacity and enthalpy values as far as large deviations from the native packing are absent and water does not penetrate in the interior of the folded regions. The thermodynamic parameter expected to be mostly affected by this assumption is the conformational entropy of the partly folded state, e.g., side chains might gain some configurational entropy if the native packing is loosened. If this is the case, our calculations will underestimate the stability of these partly folded states. Finally, by comparing the results of the calculations with experimental structural results obtained by NMR or

other techniques, it is possible to assess the extent of validity of this assumption.*

IS THE UNFOLDED STATE THE HIGHEST ENTHALPY STATE?

Figure 3 shows the calculated heat capacity for each of the 128 states of α-lactalbumin derived from the partitioning shown in Figure 2 (scheme A). For better visualization, in B the values have been normalized by the number of unfolded residues in each state. It is clear from these plots that the vast majority of states have a higher residue-normalized heat capacity than the unfolded state. There are, however, a few states that have a residue-normalized heat capacity lower than that of the unfolded state. In general, those states represent states in which the unfolded regions predominantly expose to the solvent polar surfaces but maintain significant hydrophobic contacts, i.e., hydrophobic groups that do not become exposed to water and hydrated.

As shown in the figure, there are several states with a lower normalized heat capacity than the unfolded state; in particular, states 42, 58, and 77. These states correspond to the following partly folded configurations: 0011000, 0011010, and 0011011 (in this binary representation cooperative folding units in the unfolded state are represented by a 1 and cooperative folding units in the folded state are represented by a 0; e.g., the first 0 on the left indicates that the first cooperative unit is folded).

Figure 4 shows the calculated enthalpy for each of the 128 states at different temperatures. It is clear that below 45°C the unfolded state is not the highest enthalpy state. At these temperatures, state 77 is the highest enthalpy state followed by states 58 and 42. It is also clear that as the temperature increases, the unfolded state progressively becomes the highest enthalpy state. This behavior closely resembles the behavior experimentally observed for α-lactalbumin.[5] States 77, 58, and 42 satisfy the criteria for compact denatured states since all of them preserve a significant hydrophobic core as demonstrated by their low heat capacities of 954, 802, and 685 cal $(K \cdot mol)^{-1}$, respectively, compared with the value of 1800 cal $(K \cdot mol)^{-1}$ for the unfolded state. Also, the degree of unfolding of states 77, 58, and 42 varies between 55 and 40%. All of these states, and state 77 in particular, have a heat capacity close to that observed for the acid pH compact denatured state of α-lactalbumin [980 cal $(K \cdot mol)^{-1}$].[5]

It is remarkable that the set of states with the highest enthalpies below 45°C specifies a well-defined family of states characterized by different degrees of unfolding and a clearly distinguishable structural pattern as shown in Table I. In the majority of partly folded states that satisfy the enthalpy and heat capacity criteria for the molten globule, folding units 1, 2, and 5 are folded while folding units 3 and 4 are unfolded. As seen in Figure 2, those folded units correspond primarily to the regions of the protein defined by helices A, B, and C, while the unfolded units correspond primarily to the β-sheet region.

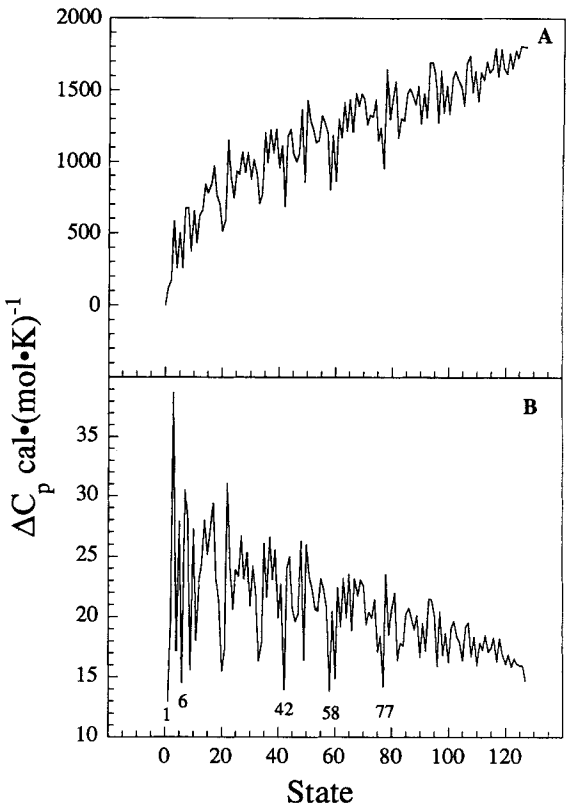

Fig. 3. **(A)** The relative heat capacity of the 128 states of α-lactalbumin. In this and all figures the thermodynamic parameters are relative to the native state. The ordering of states is proportional to the degree of unfolding of each state. **(B)** The heat capacity change in A normalized to the number of unfolded residues in each state. The five states with the lowest normalized heat capacity change (1 [0000010], 6 [0000011], 42 [0011000], 58 [0011010], 77 [0011011]) are indicated in the figure.

RESIDUAL STRUCTURE AFTER THERMAL DENATURATION

In order to assess the conditions under which partly folded states become populated, the Gibbs free energy of all species in equilibrium must be examined. According to the experimental data[5] and also to the calculations presented above, it is clear that below 45°C the unfolded state is not the highest enthalpy state of α-lactalbumin. Also, under conditions in which thermal denaturation occurs at low

*While this paper was under revision, Peng and Kim[37] published a paper on the nature of the molten globule formed by the helical domain of α-lactalbumin, suggesting that it exhibits a nativelike tertiary fold which is probably stabilized by native interactions.

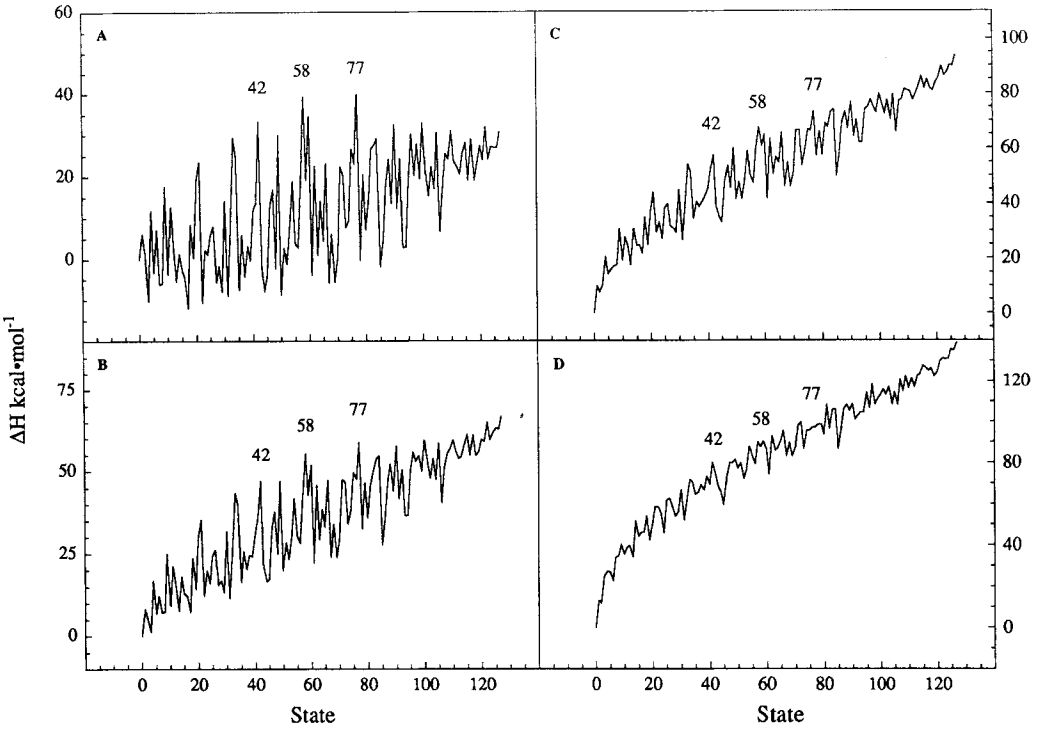

Fig. 4. The relative enthalpy of each of the 128 states of α-lactalbumin at different temperatures: **(A)** 25°C, **(B)** 45°C, **(C)** 65°C, and **(D)** 85°C. Below 45°C the unfolded state is not the highest enthalpy state. As indicated in the figure, states 42, 58, and 77 have the highest enthalpy below 45°C. At higher temperatures the unfolded state becomes the highest enthalpy state.

TABLE I. Structural Characteristics of States That Satisfy the Energetic Constraints of Molten Globule of α-Lactalbumin*

State number	Binary code†	ΔASA_{ap} (Å^2)	ΔASA_{pol} (Å^2)	$\Delta ASA_{comp,ap}$‡ (Å^2)	$\Delta ASA_{comp,pol}$‡ (Å^2)	ΔC_p§ kcal (mol·K)$^{-1}$	ΔH (25)§ kcal mol^{-1}	N_u**
21	0010000	2503.1	2074.6	720.2	552.3	0.59	23.5	34
33	0010010	3029.8	2536.6	890.9	719.1	0.7	29.4	43
42	0011000	3051.9	2647.1	541.2	449.3	0.68	33.4	49
49	0010011	3556.5	2865.5	1071.1	774.4	0.86	30.0	52
58	0011010	3578.6	3109.1	711.9	616.1	0.8	39.4	58
60	0011001	3683.0	3054.3	825.7	583.1	0.86	34.6	58
77	0011011	4105.4	3438.0	892.0	671.4	0.95	40.0	67
100	0111011	5240.9	3941.9	928.4	617.9	1.3	32.9	82
127	1111111	6772.7	4802.9	0.0	0.0	1.8	30.7	122

*States were selected on the basis of their relative enthalpy, heat capacity, and Gibbs free energy. Some states, notably 21 and 33, have an enthalpy lower than that of the unfolded state at 25°C but exhibit a low free energy during thermal denaturation (see Fig. 6).

†Each protein state is designated by a unique binary code in which a folded unit is represented by a "0" and an unfolded unit by a "1."

‡$\Delta ASA_{comp,ap}$ and $\Delta ASA_{comp,pol}$ are the apolar and polar complementary areas. The complementary area is the area located in the folded regions of a partly folded state that is buried in the native state but become exposed to the solvent in the partly folded state.

§The error estimates from the structural parametrization are ±9% for the heat capacity and ±10% for the enthalpy change at 25°C. The estimates are based on the data for the complete unfolding shown in Figure 1.

**N_u is the number of unfolded residues.

temperatures the state that becomes stable after denaturation is a compact denatured state with a higher enthalpy than the unfolded state. These results are summarized in Table I in which the enthalpy and heat capacities of the states that satisfy the criteria for the molten globule and the unfolded state are shown. Since the probability of a state is determined by the Gibbs free energy ($\Delta G = \Delta H -$

$T{\cdot}\Delta S$) it is clear that a high positive enthalpy contributes unfavorably to its stability; therefore, in order to become stable, the molten globule state must have a relative entropy higher than that of the unfolded state.

Currently, entropy changes cannot be predicted with the precision required to accurately estimate partly folded state populations. Nevertheless, in the case of α-lactalbumin this situation is alleviated by the fact that the entropy change for complete unfolding is known experimentally, leaving the entropy change of the molten globule or compact denatured state as the only adjustable parameter. It is then feasible to perform simulations in which the entropy of the molten globule state is varied systematically in order to assess the range of values under which it becomes populated. The resulting values can then be compared to those predicted from the structural analysis. The results of those simulations are summarized in Figure 5. In this figure, the heat capacity function and the population of molecules in state 77 have been represented as a function of temperature and the entropy change, $\Delta S_{77}(112)$, at the reference temperature of 112°C. State 77 has been used as an example in this figure because this state contains the consensus structure of the partly folded states in Table I. For these simulations, a $\Delta S_U(112)$ of 555 cal $(K{\cdot}mol)^{-1}$ for complete unfolding was assumed. This $\Delta S_U(112)$ value yields a transition temperature around 40°C as observed experimentally. As shown in the figure, for $\Delta S_{77}(112)$ values smaller than 370 cal $(K{\cdot}mol)^{-1}$ state 77 never becomes significantly populated and the transition is close to a two-state transition involving only the native and the unfolded state. On the other hand, for $\Delta S_{77}(112)$ values larger than 380 cal $(K{\cdot}mol)^{-1}$ state 77 is maximally populated at essentially all temperatures. At intermediate $\Delta S_{77}(112)$ values, the main temperature induced denaturation transition results in variable populations of state 77. As the temperature increases, the population of molecules in state 77 decreases and that of the unfolded state increases gradually. This behavior is similar to the one observed experimentally.[5] It is evident that for state 77 to become populated, its entropy at 112°C must be higher than 370 cal $(K{\cdot}mol)^{-1}$.

As shown in Figure 5, it is remarkable that a difference smaller than 10 cal $(K{\cdot}mol)^{-1}$ in the entropy change of the intermediate state is enough to alter significantly the character of the α-lactalbumin transition. An entropy difference of 10 cal $(K{\cdot}mol)^{-1}$ is equivalent to ~3 kcal mol^{-1} and can be elicited by changes in pH or other solvent conditions. This behavior parallels the experimental observation that at pH 5 the transition is a two-state transition with a negligible population of molecules in the compact denatured state, while at lower pH values the compact denatured state becomes significantly populated. Similar results were obtained for states 58 and 42 except that their entropy values are smaller [e.g., for state 58, $\Delta S_{58}(112)$ must be on the order of 330 cal $(K{\cdot}mol)^{-1}$ in order to become populated]. Since, at the temperatures in which it becomes populated, the compact denatured state is enthalpically unfavorable relative to the unfolded state, it is important to examine the origin of the entropic contributions that contribute to its thermodynamic stabilization.

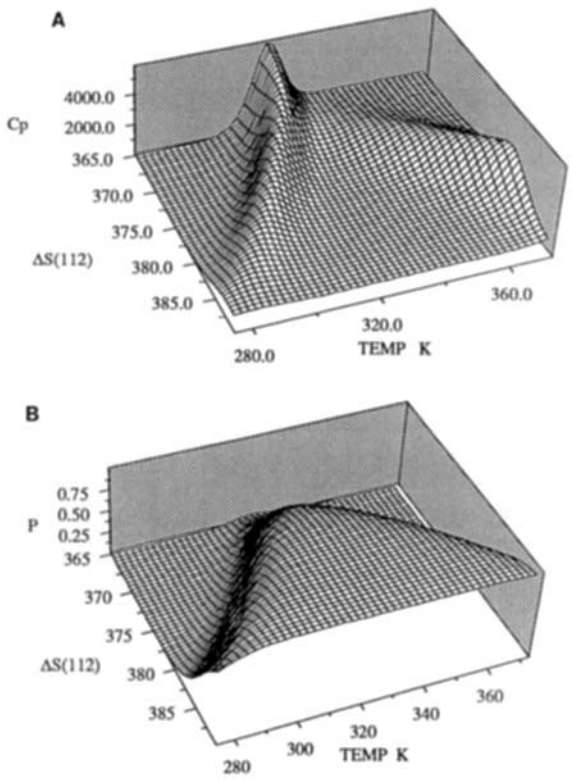

Fig. 5. The expected heat capacity function **(A)** and the population of molecules in state 77 **(B)** as a function of temperature and the entropy change, $\Delta S_7(112)$, at the reference temperature of 112°C. For $\Delta S_{77}(112)$ values smaller than 370 cal $(K{\cdot}mol)^{-1}$ state 77 never becomes significantly populated and the transition is close to a two-state transition involving only the native and the unfolded state. On the other hand, for $\Delta S_{77}(112)$ values larger than 380 cal $(K{\cdot}mol)^{-1}$ state 77 is maximally populated at essentially all temperatures. Intermediate $\Delta S_{77}(112)$ values populate the molten globule state to various levels.

CONTRIBUTIONS TO THE ENTROPY CHANGE

Several terms are known to contribute to the entropy change associated with conformational transitions in proteins. The major contributions are given by (1) "solvent"-related entropy, primarily associated with the hydrophobic effect, and (2) configurational entropy[10,23,24]:

$$\Delta S = \Delta S_{solv} + \Delta S_{conf} + \Delta S_{other}$$

The "solvent"-related and configurational entropies contribute over 95% of the total entropy of unfold-

ing.[10,11,23] Other effects include protonation or Ca^{2+} binding contributions and can be explicitly calculated from pK values, binding affinities and the experimental pH or Ca^{2+} concentration.[4] The "solvent"-related entropy, which is primarily associated with the hydrophobic effect (i.e., the unfavorable entropy associated with the exposure of apolar residues to water) and to a lesser extent to the hydration of polar groups, is given by the equation $\Delta S_{solv} = \Delta C_p \ln(T/385.15)$. It is known that at 385.15 K the experimental entropy for protein unfolding is, after correction from protonation or ligand effects, very close to the configurational entropy change or, more precisely, to the entropy change due to the increase in conformational freedom of the backbone and side chains.[10,11] Since at 385.15 K the hydrophobic contributions to ΔS are approximately zero,[23,25] it follows that the remaining terms (polar hydration, vibration, etc.) are also small or cancel each other at that temperature. Because at any temperature ΔS_{solv} is directly proportional to ΔC_p, it is clear that those states characterized by low heat capacities will exhibit the less unfavorable solvent related entropies. For example, state 77 has a solvent related entropy at 25°C of -244 cal $(K \cdot mol)^{-1}$ compared to a value of -460 cal $(K \cdot mol)^{-1}$ for the unfolded state. At that temperature, the solvent-related entropy contribution to the Gibbs free energy is about 64 kcal mol^{-1} more favorable for state 77 than for the unfolded state. This contribution more than compensates the unfavorable enthalpy difference of 10 kcal mol^{-1} at 25°C between state 77 and the unfolded state. The sum of the enthalpy and solvent entropy terms at 25°C is about 55 kcal mol^{-1} more favorable for state 77 than the unfolded state. This term should provide most of the favorable free energy necessary to overcome the unfavorable configurational entropy associated with structuring the molten globule state.

The configurational entropy change associated with structuring the molten globule state can be evaluated as follows. By definition, the completely unfolded state must have the highest configurational entropy since all intramolecular side chain interactions are disrupted and the backbone achieves a maximal conformational freedom within the constraints imposed by the presence of disulfide or other covalent links. As discussed above, $\Delta S_U(112)$ for the unfolded state is close to 555 cal $(K \cdot mol)^{-1}$ corresponding to an average 4.55 cal $(K \cdot mol)^{-1}$ per residue. Using the same per-residue value, a total of 305 cal $(K \cdot mol)^{-1}$ is expected for the unfolded regions of state 77 for example (67 amino acids in the unfolded state). To those quantities, it is necessary to add the increase in configurational side chain entropy of those amino acid residues located in the *complementary regions,* i.e., those residues located in the folded regions of partly folded states that are buried from the solvent in the native state but become exposed to the solvent in the partly folded state.[26] Analysis of the structure of state 77, for example, reveals that approximately 1564 $Å^2$ of complementary surface area corresponding to previously buried side chains become exposed to the solvent. Table I summarizes all relevant parameters necessary to perform the calculations for the states that better satisfy the thermodynamic characteristics of the molten globule state (states 21, 33, 42, 49, 58, 60, 77, and 100).

At the present time, the side chain configurational entropy gain of the amino acid residues in the complementary regions cannot be quantitated with high accuracy. However, the order of magnitude of this term can be estimated from literature values. For a series of amino acids (Ile, Leu, Met, Phe, Trp, Tyr, Val), the configurational entropy change associated with the passage of the side chain from a buried to a solvent exposed state has been estimated to range anywhere between 0.2 and 4.7 cal K^{-1} $mol\,res^{-1}$, the average being 2.8 ± 1.7 cal K^{-1} $mol\,res^{-1}$.[26,27] As a first approximation it can be assumed that this entropy term scales proportionally to the surface area that becomes exposed to the solvent by the complementary regions. If this is the case, an average value no larger than 0.04 cal $(K \cdot mol \cdot Å^2)^{-1}$ can be deduced from the structural thermodynamic database. If this scaling of the configurational entropy is assumed, then the configurational entropy contribution by the complementary regions is estimated to be on the order of 60 cal $(K \cdot mol)^{-1}$ for state 77. The complementary entropy gain will bring the total $\Delta S(112)$ to about 370 cal $(K \cdot mol)^{-1}$ for state 77. This value is very close to the value required for the stabilization of the molten globule, suggesting that standard solvent and configurational terms provide the bulk of the stabilization free energy of this state. If this is the case, the additional free energy provided by changes in solvent conditions, pH, ionic strength, etc. will be sufficient to modulate the population of molecules in the molten globule state as observed experimentally.

Finally, while our calculations assume that the folded regions conserve the packing existing in the native state, this is not necessarily so. In fact, a certain degree of packing disorder of those structural elements will additionally contribute to the configurational entropy of the molten globule state, further contributing to its stabilization.

THE GIBBS FREE ENERGY

Figure 6 shows the relative free energy of the 128 states of α-lactalbumin generated by scheme A at three different temperatures. Because of the approximations involved in the calculations, the values in the figure provide only a "coarse" estimation of the free energies and will be used only to define trends rather than to perform quantitative population cal-

culations. As shown in Figure 6, at 25°C the native state is the state with the lowest free energy and, therefore, the state that is maximally populated. At 50°C, a temperature immediately above the transition temperature, neither the native state nor the unfolded state is the lowest free energy state. At this temperature, a group of partly folded states corresponding mainly to states 21, 33, 42, 49, 58, 60, and 77 are the states with the lowest free energy. As discussed before, this group of states are all characterized by having the A, B, and C helices in the folded state and the β-sheet in the unfolded state. These results suggest that immediately after thermal denaturation a group of states characterized by having the hydrophobic core intact become the most highly populated. These states differ from each other by the state of the remaining folding units and therefore their degree of unfolding, as summarized in Table I. As the temperature increases the states with the highest degree of unfolding within this group (e.g., states 58, 77, and 100) become predominantly populated until eventually the unfolded state becomes the lowest free energy state and the most highly populated (see curve corresponding to 100°C).

The calculations presented in Figure 6 suggest that the molten globule state does not represent a unique structure but an ensemble of partly folded states all of which are characterized by a common structural feature. At low temperatures the states with the lowest degree of unfolding predominate, and at high temperatures those with the highest degree of unfolding become the predominant species. This phenomenom gives rise to a rather noncooperative and gradual thermal unfolding of the molten globule as observed by CD spectroscopy or other techniques.[1,5] In this gradual unfolding the hydrophobic core itself is the most stable part of the molten globule and the last structure to unfold. This view is similar to that advanced by Woodward[9] and consistent with the idea that the slow hydrogen exchange core is the first to fold[9] or in this case the last to unfold.

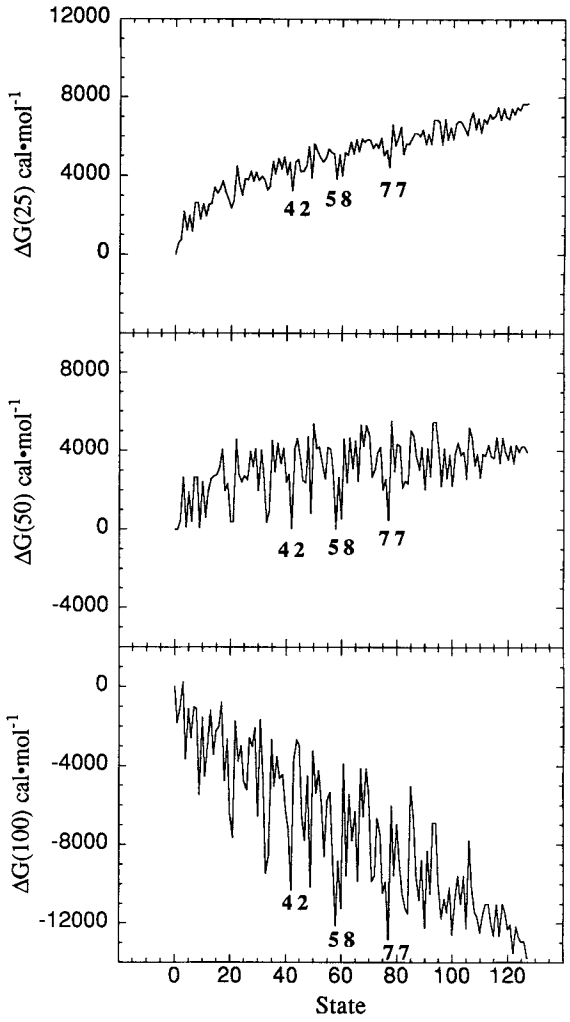

Fig. 6. The Gibbs free energy of the 128 states of α-lactalbumin at 25, 50, and 100°C. At low temperatures, the native state has the lowest free energy and is the one most significantly populated. At 50°C, a temperature immediately above the denaturation temperature, the group of partly folded states summarized in Table I becomes populated. The states in this group share common structural features and define the molten globule state. Only at higher temperatures (100°C) the unfolded state becomes significantly populated.

STRUCTURAL DATA

According to the calculations presented here, it is apparent that the main regions of α-lactalbumin that are unfolded in the molten globule or compact denatured state are those corresponding to the β-sheet, while the A, B, and C helices remain folded. These results are illustrated in Figure 7 in which the structural features of the state that contains the consensus structure of the molten globule (state 77) have been summarized. According to the NMR results[28,29] it is evident that helix B and C are folded in the acid molten globule of guinea pig α-lactalbumin. These two helices are located in folding units 2 and 5, which are folded in all the partly folded states listed in Table I. It is interesting to note that according to the hydrogen exchange experiments, some of the amide hydrogens most highly protected from exchange correspond to residues in the C helix which together with the small 3_{10}-helix forms the calcium binding site and appear to be intact and rather rigid in the A state. Not all amide hydrogens are uniformly protected in the A state; in particular the β-sheet and disordered loop amide protons exchange rapidly. The β-sheet is located in folding unit 3, which is unfolded in all the partly folded states listed in Table I. Also, the 3_{10}-helix (residues 76–82) and the two turns that form unit 4 are unfolded in agreement with the NMR data. At a more global level, the molten globule possesses a hydrophobic

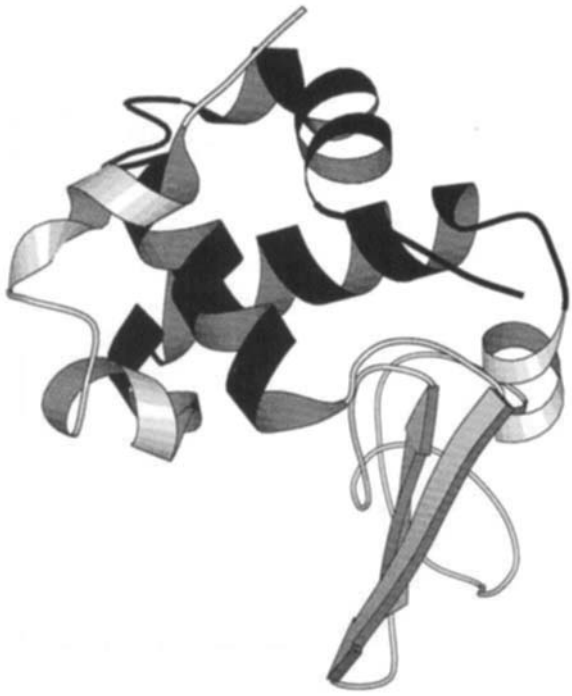

Fig. 7. The structural characteristics of state 77 (0011011) containing the consensus structure of the predicted molten globule state. In this figure, the black areas represent the regions of the protein molecule that are folded in state 77. These areas correspond to residues 1–33 and residues 83–104 (primarily helices A,B,C). The light areas correspond to regions that are unfolded in the molten globule state (primarily the β-sheet). This figure was generated with the program Molscript.[36]

core that includes 40% of all the apolar surface that is buried in the native state. On the other hand, only 28% of the native state polar contacts are maintained in this structure in agreement with the existing view that the molten globule state of α-lactalbumin is compact and has a disrupted tertiary structure.[1,4,30] NMR hydrogen exchange experiments on the kinetic folding intermediates of the structurally similar protein hen egg white lysozyme[31] indicate that in this protein the A, B, and C helices are also the earliest structural elements that become protected in pulse label experiments while the β-sheet is unfolded.

CONCLUSIONS

New studies on the energetics of the molten globule state of α-lactalbumin and other proteins[5,32,33] have revealed that many of those states have a higher enthalpy than the unfolded states at low temperatures. Despite this fact, under some conditions the molten globule state is the state that becomes populated immediately after thermal denaturation. This is possible only if this state is able to compensate its unfavorable enthalpy with a large entropic gain. Since the molten globule state is more structured than the unfolded state and is expected to have a lower configurational entropy, the net gain must originate from solvent-related entropy primarily arising from the hydrophobic effect, from protonation, and other electrostatic effects.

The results presented in this paper suggest the possibility of exploring structural features of folding intermediates by performing a structure-based thermodynamic analysis of partly folded states generated from the high resolution structure of a protein. The structural features derived using this approach are consistent with the experimental features derived by NMR. As mentioned above, our calculations assumed that the folded parts of partly folded states stay as in the native state; therefore, the search procedure is not exhaustive and necessarily limited to a subensemble of states. It is in principle feasible for some proteins to exhibit folding intermediates that do not conform to this assumption. For α-lactalbumin, the agreement observed with the experimental data suggests the absence of large deviations from the native fold in the folded regions of the molten globule state. A completely unexpected result of the calculations was that all partly folded states that satisfied the experimental thermodynamic criteria for the molten globule state belonged to the same structural family. Our initial expectation was that a diverse and heterogeneous group of states will exhibit similar thermodynamic characteristics. The results of the calculations suggest that the structural determinants of the molten globule of α-lactalbumin are well defined. This appears to be the case for other proteins as well.[34,35]

ACKNOWLEDGMENTS

We thank Dr. Bertrand Garcia-Moreno for performing the electrostatic calculations. This work was supported by grants from the National Institutes of Health (RR-04328, GM-37911, and NS-24520) and the National Science Foundation (MCB-9118687).

REFERENCES

1. Kuwajima, K. The molten globule state as a clue for understanding the folding and cooperativity of globular-protein structure. Proteins 6:87–103, 1989.
2. Dill, K.A., Shortle, D. Denatured states of proteins. Annu. Rev. Biochem. 60:795–825, 1991.
3. Ptitsyn, O. In "Protein Folding." Creighton, T.E. (eds.), New York: Freeman, 1992: 243–300.
4. Haynie, D.T., Freire, E. Structural energetics of the molten globule state. Proteins 16:115–140, 1993.
5. Griko, Y., Freire, E., Privalov, P.L. Energetics of the alpha-lactalbumin states. A calorimetric and statistical thermodynamic study. Biochemistry 33:1889–1899, 1994.
6. Dobson, C.M. Unfolded proteins, compact states and molten globules. Curr. Opinion Struct. Biol. 2:6–12, 1992.
7. Chyan, C.-L., Wormald, C., Dobson, C.M., Evans, P.A., Baum, J. Structure and stability of the molten globule state of guinea-pig alpha-lactalbumin: A hydrogen exchange study. Biochemistry 32:5681–5691, 1993.
8. Redfield, C., Smith, R.A., Dobson, C.M. Structural characterization of a highly-ordered "molten globule" at low pH. Struct. Biol. 1:23–29, 1994.
9. Woodward, C. Is the slow-exchange core the protein folding core? TIBS 18:359–360, 1993.
10. Murphy, K.P., Bhakuni, V., Xie, D., Freire, E. The molec-

ular basis of cooperativity in protein folding III: Identification of cooperative folding units and folding intermediates. J. Mol. Biol. 227:293–306, 1992.
11. Murphy, K.P., Freire, E. Thermodynamics of structural stability and cooperative folding behavior in proteins. Adv. Protein Chem. 43:313–361, 1992.
12. Spolar, R.S., Livingstone, J.R., Record, M.T., Jr. Use of liquid hydrocarbon and amide transfer data to estimate contributions to thermodynamic functions of protein folding from the removal of nonpolar and polar surface from water. Biochemistry 31:3947–3955, 1992.
13. Murphy, K.P., Xie, D., Garcia, K.C., Amzel, L.M., Freire, E. Structural energetics of peptide recognition: Angiotensin II/antibody binding. Proteins 15:113–120, 1993.
14. Thompson, K., Vinson, C., Freire, E. Thermodynamic characterization of the structural stability of the coiled-coil region of the bZIP transcription factor GCN4. Biochemistry 32:5491–5496, 1993.
15. Freire, E. Structural thermodynamics: Prediction of protein stability and protein binding affinities. Arch. Biochem. Biophys. 303:181–184, 1993.
16. Lee, B., Richards, F.M. The interpretation of protein structures: Estimation of static accessibility. J. Mol. Biol. 55: 379–400, 1971.
17. Privalov, P.L., Makhatadze, G.I. Contribution of hydration to protein folding thermodynamics. II. The entropy and Gibbs energy of hydration. J. Mol. Biol. 232:660–679, 1993.
18. Acharya, K.R., Stuart, D.I., Walker, N.P.C., Lewis, M., Phillips, D.C. Refined structure of baboon alpha-lactalbumin at 1.7 Å resolution. J. Mol. Biol. 208:99–127, 1989.
19. Hughson, F.M., Wright, P.E., Baldwin, R.L. Structural characterization of a partly folded apomyoglobin intermediate. Science 249:1544–1548, 1990.
20. Weissman, J.S., Kim, P.S. Reexamination of the folding of BPTI: Predominance of native intermediates. Science 253: 1386–1393, 1991.
21. Jennings, P.A., Wright, P.E. Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin. Science 262:892–896, 1993.
22. Poland, D., Scheraga, H. "Theory of Helix-Coil Transitions in Biopolymers." New York: Academic Press, 1970.
23. Privalov, P.L., Gill, S.J. Stability of protein structure and hydrophobic interaction. Adv. Protein Chem. 39:191–234, 1988.
24. Dill, K.A. Dominant forces in protein folding. Biochemistry 29:7133–7155, 1990.
25. Baldwin, R.L. Temperature dependence of the hydrophobic interaction in protein folding. Proc. Natl. Acad. Sci. U.S.A. 83:8069–8072, 1986.
26. Freire, E., Haynie, D.T., Xie, D. Molecular basis of cooperativity in protein folding. IV. CORE: A general cooperative folding model. Proteins 17:111–123, 1993.
27. Creamer, T.P., Rose, G.D. Side-chain entropy opposes α-helix formation but rationalizes experimentally determined helix-forming propensities. Proc. Natl. Acad. Sci. U.S.A. 89:5937–5941, 1992.
28. Baum, J., Dobson, C.M., Evans, P.A., Hanley, C. Characterization of a partly folded protein by NMR methods: Studies of the molten globule state of guinea-pig α-lactalbumin. Biochemistry 28:7–13, 1989.
29. Alexandrescu, A.T., Evans, P.A., Pitkeathly, M., Baum, J., Dobson, C.M. Structure and dynamics of the acid-denatured molten globule state of alpha-lactalbumin: A two-dimensional NMR study. Biochemistry 32:1707–1718, 1993.
30. Xie, D., Bhakuni, V., Freire, E. Calorimetric determination of the energetics of the molten globule intermediate in protein folding: Apo-α-lactalbumin. Biochemistry 30: 10673–10678, 1991.
31. Radford, S.E., Dobson, C.M., Evans, P.A. The folding of hen lysozyme involves partially structured intermediates and multiple pathways. Nature (London) 358:302–307, 1992.
32. Kuroda, Y., Kidokoro, S.-I., Wada, A. Thermodynamic characterization of cytochrome c at low pH. Observation of the molten globule state and of the cold denaturation process. J. Mol. Biol. 223:1139–1153, 1992.
33. Murphy, K.P., Robertson, A.D., Meadow, N.D., Roseman, S., Freire, E. Residual structure in a thermally denatured β-sheet protein. In preparation, 1994.
34. Freire, E. and Xie, D. Thermodynamic Prediction of Structural Determinants of the Molten Globule State of Bacnare. Biophys. Chem. In Press: 1994.
35. Xie, D., Freire, E. Structure based prediction of protein folding intermediates. J. Mol. Biol. In Press.
36. Kraulis, P.J. MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. J. Appl. Crystallogr. 24:946–950, 1991.
37. Peng, Z-Y, Kim, P.S. A protein dissection study of a molten globule. Biochemistry 33:2136–2141, 1994.