
By Ernesto Freire

Introduction

In 1978 Freire and Biltonen realized that the heat capacity function measured by differential scanning calorimetry (DSC) contained enough information to calculate the folding/unfolding partition function of a protein, and that this partition function could be used to evaluate the number of states that become populated during the transition and the thermodynamic parameters associated with those states. It has since been shown that the thermodynamic information obtained calorimetrically can be linked directly to the structure of a protein in a quantitative fashion. In fact, this development has opened the doors to a direct structural deconvolution of the heat capacity function. The structural thermodynamic approach has permitted the development of an accurate structural parametrization of protein energetics and the possibility of evaluating the folding/unfolding partition function from the crystallographic or nuclear magnetic resonance (NMR) solution structure of a protein. In this chapter the linkage between the thermodynamics and structure through the partition function is discussed.

Folding/Unfolding Partition Function

The most fundamental quantity required to account for the stability and folding behavior of monomeric proteins is the partition function, $Q$, defined as the sum of the statistical weights of all the states accessible to a protein:

$$Q = \sum_{i=0}^{N} \exp\left(\frac{-\Delta G_i}{RT}\right)$$

where the statistical weights or Boltzmann exponents \([\exp(-\Delta G_i/RT)]\) are defined in terms of the Gibbs free energy \(\Delta G_i\) for each state, \(R\) is the gas constant, and \(T\) is the absolute temperature. The Gibbs free energy of each state is given by the standard thermodynamic relationship:

\[
\Delta G_i = \Delta H_i(T_R) + \Delta C_{p,i}(T - T_R) - T[\Delta S_i(T_R) + \Delta C_{p,i} \ln(T/T_R)]
\]  

(2)

where \(\Delta H_i(T_R)\) and \(\Delta S_i(T_R)\) are the relative enthalpy and entropy of state \(i\) at the reference temperature \(T_R\), and \(\Delta C_{p,i}\) is the relative heat capacity of that state. For convenience, the native state is chosen as the reference state to express all relative thermodynamic parameters.

Equation (1) can be written as

\[
Q = 1 + \sum_{i=1}^{N-1} \exp(-\Delta G_i/RT) + \exp(-\Delta G_N/RT)
\]  

(3)

where the terms under the summation sign include all intermediates that become populated during the transition. The first and last terms are the statistical weights of the native and unfolded states, respectively. In principle the number of terms within the summation is astronomical; in practice, however, folding/unfolding transitions are highly cooperative, and most of those terms are negligibly small. In fact, under many experimental conditions the folding/unfolding transition is well approximated by the two-state partition function \([Q = 1 + \exp(-\Delta G_N/RT)]\). Under certain experimental conditions (e.g., extreme pH values, mild concentration of denaturants, extreme ionic strengths) or after certain mutations, some proteins exhibit significant populations of stable partly folded intermediates.

All thermodynamic parameters can be expressed in terms of the partition function. For example, the population of molecules in state \(i\) \((P_i)\) is equal to the ratio of the statistical weight of that state over the sum of the statistical weights of all the states:

\[
P_i = \frac{\exp(-\Delta G_i/RT)}{\left[1 + \sum_{i=1}^{N} \exp(-\Delta G_i/RT)\right]}
\]  

(4)

Fig. 1. Molar heat capacity function \( (C_p) \) obtained by differential scanning calorimetry. The molar heat capacities of the unfolded \( (C_{p,u}) \) and native \( (C_{p,n}) \) states are indicated by the dashed lines. The difference between these values define the heat capacity change for the transition \( (\Delta C_p = C_{p,u} - C_{p,n}) \). The average excess heat capacity function \( (\langle \Delta C_p \rangle) \) is obtained by subtracting the heat capacity of the native state from the molar heat capacity function \( (\langle \Delta C_p \rangle = C_p - C_{p,n}) \).

The average excess enthalpy function \( (\langle \Delta H \rangle) \) is the sum of the relative enthalpy contributions of all the states that become populated during the transition\(^1\):

\[
\langle \Delta H \rangle = \sum_{i=0}^{N} P_i \Delta H_i \tag{5a}
\]

\[
= RT^2 (\partial \ln Q / \partial T) \tag{5b}
\]

The excess heat capacity function measured by DSC, \( \langle \Delta C_p \rangle \), is equal to the temperature derivative of \( \langle \Delta H \rangle \) at constant pressure:

\[
\langle \Delta C_p \rangle = (\partial \langle \Delta H \rangle / \partial T)_p \tag{6a}
\]

\[
\langle \Delta C_p \rangle = \sum_{i=1}^{N} \Delta H_i (\partial P_i / \partial T) + \sum_{i=1}^{N} P_i \Delta C_{p,i} \tag{6b}
\]

\[
= \langle \Delta C_{p,u} \rangle + \langle \Delta C_{p,bl} \rangle \tag{6c}
\]

The first term on the right-hand side \( (\langle \Delta C_{p,u} \rangle) \) is the transition excess heat capacity function and defines the characteristic transition peak(s) in the heat capacity function (Fig. 1).\(^1\) The second term on the right-hand side

\( \langle (\Delta C_{P, b}) \rangle \) defines the "S-shape" shift in baseline usually associated with protein unfolding or other transitions characterized by positive changes in \( \Delta C_P \).\(^{14}\)

The above treatment is completely general for a monomeric system in equilibrium and does not depend on \textit{a priori} assumptions regarding the mechanism of the folding/unfolding transition or the number of intermediate states. It must be noted that three thermodynamic parameters \([\Delta H_i(T_R), \Delta S_i(T_R), \text{and } \Delta C_{P,i}]\) are required to specify each state in the partition function. Because one of those states is chosen as the reference state (the native state) only \(3N\) parameters are necessary to describe a \((N + 1)\) state system.

Evaluation of the partition function requires the identification and enumeration of the relevant folding states (folded, unfolded, and partially folded states) of the protein and their Gibbs free energies. Several years ago, Freire and Biltonen\(^1\) showed that scanning calorimetry data could be used to evaluate experimentally the protein folding/unfolding partition function. More recently, the development of a structural parametrization of the protein folding energetics\(^4\) has set the basis for the development of algorithms aimed at evaluating the partition function from the crystallographic or NMR solution structure of a protein.\(^2,3,15\)

As illustrated in Diagram 1 above, there are two ways of obtaining information regarding the partition function: (1) from calorimetric data and (2) from the crystallographic or NMR solution structure of a protein. In this chapter we discuss those two approaches.

Experimental Determination of Thermodynamic Parameters

The thermodynamic parameters \((\Delta H, \Delta S, \Delta C_P)\) that define the Gibbs free energy difference between two protein states can be measured by DSC.

Excess Heat Capacity Function

Differential scanning calorimetry measures the heat capacity of a protein as a function of temperature. If the protein undergoes a temperature-induced transition, its heat capacity will exhibit an anomaly centered at a characteristic temperature called the transition temperature ($T_m$). Within the transition region, the heat capacity function contains contributions from all the states that become populated during the transition as well as the excess contributions arising from the existence of transition-related enhanced enthalpy fluctuations. These enhanced fluctuations give rise to the characteristic peak or peaks associated with thermally induced transitions.\(^{16,17}\)

The excess heat capacity function is obtained by subtracting the heat capacity of the native state from the measured heat capacity function\(^1\):

$$\langle \Delta C_p \rangle = \langle C_p \rangle - C_{p,n} \quad (7)$$

Figure 1 illustrates the procedure required to estimate $\langle \Delta C_p \rangle$ from the experimental data. As indicated in Fig. 1, this procedure involves extrapolation of $C_{p,n}$ to those temperature regions in which the native state is no longer the most significantly populated. The subtraction of $C_{p,n}$ from $\langle C_p \rangle$ implicitly selects the native state as the reference state.

Overall Thermodynamic Parameters

The total enthalpy ($\Delta H$), entropy ($\Delta S$), and heat capacity ($\Delta C_p$) changes between the unfolded and native states are the most important overall thermodynamic parameters determined by DSC since they define the Gibbs free energy of stabilization of the native state [see Eq. (2)]. These thermodynamic parameters are state functions, that is, their value depends only on the nature of the unfolded and the native states and not on the specific transition pathway or the presence of partly folded intermediates. From a practical point of view, $\Delta H$, $\Delta S$, and $\Delta C_p$ are independent of the shape of the measured heat capacity function. The heat capacity change is defined by the value of $\langle \Delta C_p \rangle$ after completion of the transition as indicated in Fig. 1. The enthalpy change is the area under the transition excess heat capacity function ($\langle \Delta C_{p,\text{tr}} \rangle$)

$$\Delta H = \int_{T_0}^{T_f} \langle \Delta C_{p,\text{tr}} \rangle \, dT \quad (8)$$


where the limits of integration are defined by the onset and completion temperatures of the transition (i.e., the temperatures at which essentially all molecules are in the initial and final states, respectively). The entropy change is simply evaluated by means of

\[ \Delta S = \int_{T_o}^{T_f} (\Delta C_{p, tr}) d \ln T \]  

Both \( \Delta H \) and \( \Delta S \), as defined by Eqs. (8) and (9), are referred to the transition temperature, \( T_m \), that is, \( \Delta H = \Delta H(T_m) \) and \( \Delta S = \Delta S(T_m) \).

Traditionally, the evaluation of the two-state or multistate character of a transition has been done by comparing the calorimetrically determined enthalpy [Eq. (8)] with the van’t Hoff enthalpy which calorimetrically is usually evaluated with the following formula:

\[ \Delta H_{vH} = \frac{4RT_m^2 \langle \Delta C_{p, tr} \rangle_{T_m}}{\Delta H} \]  

where \( T_m \) is the temperature at which \( \langle \Delta C_{p, tr} \rangle \) is maximal and \( \langle \Delta C_{p, tr} \rangle_{T_m} \) is the value of \( \langle \Delta C_{p, tr} \rangle \) at \( T_m \). Some authors use the temperature of half-completion of the transition to evaluate \( \Delta H_{vH} \). Usually a \( \Delta H_{vH}/\Delta H \) ratio equal to one is taken as an indication that the transition under study conforms to the two-state mechanism; a \( \Delta H_{vH}/\Delta H \) ratio smaller than one as an indication of the presence of significantly populated intermediates; and a \( \Delta H_{vH}/\Delta H \) ratio greater than one as an indication of intermolecular cooperation.

Even though it is widely used, the \( \Delta H_{vH}/\Delta H \) ratio does not provide an accurate measurement of the character of a transition. For example, it has been concluded that the unfolding of small single-domain globular proteins obeys very closely the two-state model based on this criterion. Although this conclusion is correct for some proteins under specific conditions, it is certainly not true for all proteins. Also, proteins that display two-state behavior under certain solvent conditions do not behave that way under other conditions. It has been shown, for example, that myoglobin, cytochrome c, and apo-\( \alpha \)-lactalbumin do not undergo two-state transitions under all conditions.

The inaccuracy of the \( \Delta H_{vH}/\Delta H \) ratio to account for the two-state behavior of a protein can be illustrated with a close examination of the

experimental values obtained for globular proteins. For example, Privalov and Khechinashvili \cite{21} obtained an average $\Delta H_{vH}/\Delta H$ ratio of 0.95, the extreme values being 0.9 and 1.02. It is significant that the average $\Delta H_{vH}/\Delta H$ ratio is less than one, suggesting that the folding/unfolding reactions are not really two-state transitions and that they exhibit some population of intermediates. Based on the small numerical deviation of the $\Delta H_{vH}/\Delta H$ ratio from unity, it would be erroneous to generalize that the population of intermediates must be small in all those cases.\cite{19} In fact, intermediate populations as high as 25% are consistent with the experimental $\Delta H_{vH}/\Delta H$ ratios obtained for small globular proteins.

One of the reasons for the inaccuracy of the $\Delta H_{vH}/\Delta H$ ratio is that it utilizes only a single point from the entire calorimetric scan. Also, and perhaps more importantly, its numerical value is not directly related to a specific population of intermediate states. Thus, it is possible for a transition with a significant population of intermediates to have a $\Delta H_{vH}/\Delta H$ ratio close to one and for a transition with a small intermediate population to have a $\Delta H_{vH}/\Delta H$ ratio significantly different than one. Of particular importance are intermediate states that are either enthalpically close to the native state or close to the unfolded state. The presence of relatively high populations of those intermediates within the transition region will not cause significant deviations from unity in the $\Delta H_{vH}/\Delta H$ ratio. It is evident that a rigorous analysis of the heat capacity function for systems involving more than two states requires a complete statistical thermodynamic analysis and not a simple guess based on the $\Delta H_{vH}/\Delta H$ ratio.

**Experimental Evaluation of Partition Function**

By definition, $\langle \Delta H \rangle$ is the cumulative integral of $\langle \Delta C_p \rangle$:

$$\langle \Delta H \rangle = \int_{T_o}^{T} \langle \Delta C_p \rangle \, dT \quad (11)$$

where $T_o$ is a temperature at which the protein is in the native state (Fig. 1). $\langle \Delta H \rangle$ plays a central role in the statistical thermodynamic analysis of DSC data because it provides a direct link between the experiment and the folding/unfolding partition function. Freire and Biltonen\cite{1} first realized that, by rewriting Eq. (5b) in integral form, DSC could provide a direct numerical access to the folding/unfolding partition function:

\[ \ln Q = \int_{T_0}^{T} \frac{\langle \Delta H \rangle}{RT^2} dT \]  
(12)

\[ \ln Q = \int_{T_0}^{T} \frac{1}{RT^2} \left( \int_{T_0}^{T} \langle \Delta C_p \rangle dT \right) dT \]  
(13)

Eqs. (12) and (13) provide a rigorous foundation to the deconvolution theory of the excess heat capacity function since they establish a mathematical linkage between the experimental data and the most fundamental function in statistical thermodynamics: the partition function.

Deconvolution of Excess Heat Capacity Function

The main goal of the deconvolution analysis of the heat capacity function is the determination of the number of states that become populated during thermal unfolding and the thermodynamic parameters for each of those states. Throughout the years the deconvolution algorithms have been perfected in different ways (see, e.g., Refs. 1 and 22–24).

In general, the deconvolution procedure consists of two parts: (1) an initial estimation of the number of states and associated thermodynamic parameters and (2) a global optimization of the parameters using nonlinear least squares. The most accurate method for initial parameter estimation involves the recursive deconvolution algorithm\(^1\) since it yields parameter estimates that are already close to the final convergence values even for strongly overlapped transitions.

The recursive deconvolution algorithm is based on the following general relationship:

\[ d \ln (Q - 1)/dT = \Delta H_1 + \langle \Delta H_1 \rangle \]  
(14)

where \( \Delta H_1 \) is the enthalpy difference between the first intermediate and the native state, and \( \langle \Delta H_1 \rangle \) is the average excess enthalpy function that would exist if the first intermediate were the lowest enthalpy state and the native state did not exist.\(^1\) Equation (14) allows determination of the enthalpy difference between the first intermediate and the native state without any assumptions regarding the total number of states in the transition. \( \langle \Delta H_1 \rangle \) is then used to define a new partition function in which the first intermediate becomes the reference state. A second application of Eq. (14) yields the enthalpy for the second intermediate and a new partition


function. This procedure is repeated until the entire excess enthalpy is accounted for. Multiple passes and intermediate nonlinear least squares optimization are performed at each step in the deconvolution procedure in order to minimize error propagation.

After initial parameter estimation, a global nonlinear least squares optimization is performed in order to determine the best set of parameter values. The nonlinear least squares fit of \( \langle \Delta C_p \rangle \) must be performed using Eq. (6b), which is the exact equation for a folding/unfolding transition exhibiting an arbitrary number of states. In the past some authors have only used the first term of Eq. (6b)\(^2\); however, this practice yields incorrect results for transitions with significant \( \Delta C_p \) values.\(^2\) After convergence, the nonlinear least squares procedure returns the best set of parameters that minimizes the sum of squared residuals (SSR) between the calculated and experimental values [SSR = \( \sum (\langle \Delta C_p \rangle_{\text{calculated}} - \langle \Delta C_p \rangle_{\text{experimental}})^2 \)]. After completion of the nonlinear least squares optimization the goodness of the fit must be evaluated in terms of the standard deviation of the fit and by performing an analysis of residuals.\(^2\) Because there are three extra fitting parameters (\( \Delta H, \Delta S, \Delta C_p \)) for each additional state that is included in the analysis, it is expected that the goodness of the fit will increase if the number of states is increased. It is necessary, then, to evaluate whether the increase in the goodness of the fit actually reflects the existence of an additional state or is merely due to the larger number of parameters used in the analysis. This can be done by performing an F-test statistic as described by Draper and Smith.\(^2\)

Structural Thermodynamics

The thermodynamic parameters (\( \Delta H, \Delta S, \Delta C_p \)) that define the Gibbs free energy difference between two protein states can be parametrized in terms of the structures of those states.

**Heat Capacity of Unfolded State**

The absolute heat capacity of a protein in the unfolded state (\( C_{p,u} \)) can be calculated with high accuracy from the amino acid sequence. If all the constituent groups are exposed to water, then \( C_{p,u} \) obeys simple additivity Rules and can be expressed in terms of the individual contributions from the amino acid side chains and the peptide backbone:

\[
C_{p,u} = \left( \sum_{i=1}^{20} n_i C_{p,i} \right) + (N_{AA} - 1) C_{p,bb} + C_{p,NH_2} + C_{p,COOH} \quad (15)
\]

where the index in the summation refers to the 20 amino acids, $n_i$ is the number of amino acids of type $i$ and $C_{p,i}$ the molar heat capacity of its side chain; $N_{AA}$ is the number of amino acids in the protein; and $C_{p,bb}$ is the heat capacity of a peptide backbone unit (—CHCONH—), $C_{p,NH_2}$ the heat capacity of the amino terminus, and $C_{p,COOH}$ the heat capacity of the carboxyl terminus. Table I summarizes the parameters necessary to estimate $C_{p,u}$ from the amino acid sequence of a protein. The parameters in Table I were obtained by a polynomial fit of the data published by

<table>
<thead>
<tr>
<th>Contributions from</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Side-chain amino acid</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td>177.957</td>
<td>-0.450</td>
<td>0.000153</td>
<td></td>
</tr>
<tr>
<td>Arg</td>
<td>299.558</td>
<td>0.627</td>
<td>-0.00267</td>
<td></td>
</tr>
<tr>
<td>Asn</td>
<td>68.395</td>
<td>0.884</td>
<td>-0.00160</td>
<td></td>
</tr>
<tr>
<td>Asp</td>
<td>68.731</td>
<td>0.8206</td>
<td>-0.00107</td>
<td></td>
</tr>
<tr>
<td>Cys</td>
<td>222.58</td>
<td>0.636</td>
<td>-0.0017</td>
<td></td>
</tr>
<tr>
<td>Gln</td>
<td>165.183</td>
<td>0.6360</td>
<td>-0.0017</td>
<td></td>
</tr>
<tr>
<td>Glu</td>
<td>164.742</td>
<td>0.636</td>
<td>-0.00168</td>
<td></td>
</tr>
<tr>
<td>Gly</td>
<td>83.505</td>
<td>-0.226</td>
<td>-9.54 x 10^{-5}</td>
<td></td>
</tr>
<tr>
<td>His</td>
<td>212.367</td>
<td>-1.699</td>
<td>0.0230</td>
<td>-8.36 x 10^{-5}</td>
</tr>
<tr>
<td>Ile</td>
<td>408.045</td>
<td>-0.233</td>
<td>0.000115</td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>386.787</td>
<td>-0.198</td>
<td>0.000225</td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td>316.945</td>
<td>-1.365</td>
<td>0.0188</td>
<td></td>
</tr>
<tr>
<td>Met</td>
<td>200.523</td>
<td>-1.0456</td>
<td>0.00487</td>
<td></td>
</tr>
<tr>
<td>Phe</td>
<td>398.560</td>
<td>-0.6340</td>
<td>0.00130</td>
<td></td>
</tr>
<tr>
<td>Pro</td>
<td>218.452</td>
<td>-1.623</td>
<td>0.0075</td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td>74.736</td>
<td>0.233</td>
<td>-0.000108</td>
<td></td>
</tr>
<tr>
<td>Thr</td>
<td>196.165</td>
<td>-0.571</td>
<td>0.00589</td>
<td></td>
</tr>
<tr>
<td>Trp</td>
<td>474.060</td>
<td>-0.634</td>
<td>0.00130</td>
<td></td>
</tr>
<tr>
<td>Tyr</td>
<td>314.312</td>
<td>-0.716</td>
<td>0.00785</td>
<td>-2.2 x 10^{-5}</td>
</tr>
<tr>
<td>Val</td>
<td>325.226</td>
<td>-0.361</td>
<td>-0.000586</td>
<td></td>
</tr>
<tr>
<td>Peptide backbone unit (—CHCONH—)</td>
<td>1.273</td>
<td>0.613</td>
<td>-0.00286</td>
<td></td>
</tr>
<tr>
<td>Amino Terminus (NH_2)</td>
<td>-192.630</td>
<td>6.178</td>
<td>-0.0666</td>
<td>0.000157</td>
</tr>
<tr>
<td>Carboxyl terminus (COOH)</td>
<td>1.273</td>
<td>0.613</td>
<td>-0.00286</td>
<td></td>
</tr>
</tbody>
</table>

The coefficients were obtained by a polynomial least squares fit of the data published by Privalov and Makhatadze26 (E. Freire and K. P. Murphy, unpublished results, 1993). The form of the equation is

$$C_{p,u}(T) = A + BT + CT^2 + DT^3$$

where $C_p$ is in cal K$^{-1}$ mol$^{-1}$ and $T$ in °C.
Heat Capacity Changes

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 \( \Delta ASA_{pol} \) and apolar \( \Delta ASA_{ap} \) solvent accessible surface areas between those states\(^{4,27-31}\):

\[
\Delta C_p = \Delta C^{\circ}_{p,ap} \Delta ASA_{ap} + \Delta C^{\circ}_{p,pol} \Delta ASA_{pol}
\]

(16)

where \( \Delta C^{\circ}_{p,ap} \) \([0.45 \pm 0.02 \text{ cal K}^{-1} (\text{mol-Å}^2)^{-1}] \) and \( \Delta C^{\circ}_{p,pol} \) \([-0.26 \pm 0.03 \text{ cal K}^{-1} (\text{mol-Å}^2)^{-1}] \) are the elementary apolar and polar contributions to the total heat capacity increment, and \( \Delta ASA_{ap} \) and \( \Delta ASA_{pol} \) are the differences between the solvent accessible apolar and polar surface areas of the two states. This parametrization has been shown to predict the \( \Delta C_p \) of protein folding/unfolding with an average deviation better than \( \pm 9\% \).\(^{4,6,30}\)

Heat Capacity of Native State

The heat capacity of the native state, \( C_{p,n} \), can be written in terms of Eqs. (15) and (16) as

\[
C_{pn} = C_{p,u} - \Delta C^{\circ}_{p,ap} \Delta ASA_{ap} - \Delta C^{\circ}_{p,pol} \Delta ASA_{pol}
\]

(17)

The above approach accurately predicts the heat capacity values obtained for the native state of globular proteins and the magnitude of the heat capacity change associated with the complete unfolding of the protein. The heat capacity values provide a means to evaluate the degree of unfolding of a protein by comparing experimentally determined values to those calculated with Eqs. (14)–(16). In conjunction with the measured changes in \( \Delta H \) and \( \Delta S \) on unfolding, the heat capacity values provide a rather complete assessment of the degree of unfolding of a protein (see below).

Privalov and Makhatadze\(^{32}\) have assumed that, unlike the heat capacity of the unfolded state, the heat capacity of the native state is a linear

---

function of temperature and that, consequently, $\Delta C_p$ decreases at higher temperatures. It must be noticed, however, that in some cases in which the heat capacity of the native state has been measured accurately such linearity has not been observed (see, e.g., Ref. 29). In any case, the resulting temperature dependence of $\Delta C_p$ is usually not very large. In fact, in most cases in which the temperature dependence of the enthalpy change has been measured over a wide temperature interval, no evidence of a diminished slope at high temperatures has been observed.\(^{17,30,31}\) For most practical purposes, the enthalpy change exhibits a linear temperature dependence within the temperature range of interest (0$^\circ$–100$^\circ$).

**Enthalpy Change**

The enthalpy change associated with protein unfolding is due to the disruption of intramolecular interactions (van der Waals, hydrogen bonding, etc.), the concomitant formation of interactions with water owing to hydration of the groups that become unfolded, and to a lesser extent the protonation or ionization of histidyl and carboxyl groups, the release of specific ligands or prosthetic groups, etc. In the absence or after correction for protonation or ligand binding contributions, the enthalpy change can be written as a linear combination of the changes in apolar and polar accessible surface areas:

$$
\Delta H(T) = a(T)\Delta ASA_{ap} + b(T)\Delta ASA_{pol} \quad (18)
$$

$$
\Delta H(T) = a(T_R)\Delta ASA_{ap} + b(T_R)\Delta ASA_{pol} + \Delta C_p(T - T_R) \quad (18a)
$$

where $T_R$ is an appropriately chosen reference temperature. In previous papers, we used 100$^\circ$ 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. However, we have found that a better reference temperature is 60$^\circ$, which corresponds to the median transition temperature for the proteins in the database.\(^6\) This choice minimizes extrapolation errors arising from uncertainties in $\Delta C_p$ and the possibility that $\Delta C_p$ for some proteins might not be constant and that it decreases at higher temperatures as suggested by Privalov and Makhatadze.\(^{33}\) At 60$^\circ$ the enthalpy change can be written as:

$$
\Delta H(60) = 31.4\Delta ASA_{pol} - 8.44\Delta ASA_{ap} \quad (19)
$$

Equation (19) predicts the experimental enthalpy at 60$^\circ$ within $\pm6\%$ for the protein thermodynamic database, that is, the set of proteins for

---

which high-resolution structure and thermodynamic information are available. At this writing the following proteins are included in the protein thermodynamic database: cytochrome c, carbonic anhydrase (carbonate dehydratase), chymotrypsin, α-lactalbumin, lysozyme, myoglobin, staphylococcal nuclease, papain, parvalbumin, pepsinogen, ribonuclease A, ribonuclease T1, trypsin, protein G, and basic pancreatic trypsin inhibitor (BPTI). Figure 2 illustrates the differences between calculated and experimental enthalpy and heat capacity changes for protein unfolding.

Under most conditions, Eqs. (18) and (19) 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. (18) 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$^{-1}$.

Entropy Change

The entropy change for the unfolding of globular proteins exhibits a convergence temperature, that is, a temperature at which the residue normalized unfolding entropies assume a similar value.$^{34}$ It has been shown that the entropy convergence occurs at $112^\circ$ and that at this temperature the entropy change is close to the change in entropy due to the increase in the conformational degrees of freedom of the protein backbone and side chains.$^4$ Previously Baldwin$^{35}$ has shown that at $112^\circ$ the entropy associated with the dissolution of hydrophobic groups is very close to zero. These results imply that the remaining entropy terms (polar hydration, vibrational, etc.) are either negligible or cancel one another. Thus, at any temperature the unfolding entropy change can be written as

$$\Delta S(T) = \Delta S_{\text{conf}} + \Delta C_p \ln(T/385.15)$$

(20)

where $\Delta S_{\text{conf}}$ is the conformational or configurational entropy change.

It has been proposed$^3$ that the configurational entropy can be written in terms of a minimum of three different types of contributions: (1) $\Delta S_{\text{bu-ex}}$, the entropy change associated with the transfer of a side chain that is buried in the interior of the protein to its surface; (2) $\Delta S_{\text{ex-su}}$, the entropy


Fig. 2. Comparison between the experimental and calculated heat capacity, $\Delta C_p$, and enthalpy, $\Delta H(60)$, changes for unfolding of the proteins in the thermodynamic database. (Reprinted from Xie and Freire.)
change gained by a surface exposed side chain when the peptide backbone unfolds; and (3) $\Delta S_{bb}$, the entropy change gained by the backbone itself on unfolding. The total entropy change can be written as

$$\Delta S_{\text{conf}} = \sum_i \Delta S_{\text{ex} \rightarrow u,i} + \sum_j \Delta S_{\text{bu} \rightarrow \text{ex},j} + \Delta S_{bb}$$  \hspace{1cm} (21)$$

where the summation $i$ runs over all amino acid side chains and the summation $j$ runs only over those amino acids that are buried.

Creamer and Rose\textsuperscript{36} estimated side-chain contributions to the configurational entropy change of helix-to-coil transitions in isolated peptides using Monte Carlo simulations. These authors estimated the equivalent of $\Delta S_{\text{ex} \rightarrow u}$ for seven amino acids (Ile, Leu, Met, Phe, Trp, Tyr, Val), finding that it ranges from about 0.14 cal K$^{-1}$ mol$^{-1}$ for leucine to 1.28 cal K$^{-1}$ mol$^{-1}$ for valine and averages 0.66 $\pm$ 0.46 cal K$^{-1}$ mol$^{-1}$. Blaber et al.\textsuperscript{37} have also estimated $\Delta S_{\text{ex} \rightarrow u}$ theoretically from an analysis of side-chain conformations in 100 crystallographic structures. These authors obtained a mean value of 0.56 $\pm$ 0.48 cal K$^{-1}$ mol$^{-1}$ for 19 amino acids, which is close to the value obtained by Creamer and Rose\textsuperscript{36} and also similar to the value of 0.5 cal K$^{-1}$ mol$^{-1}$ obtained much earlier by Nemethy et al.\textsuperscript{38} Creamer and Rose\textsuperscript{36} also calculated the configuration entropy of side chains in the $\alpha$ helix, which ranges between 0.2 and 4.7 cal K$^{-1}$ mol$^{-1}$, the average being 2.8 $\pm$ 1.7 cal K$^{-1}$ mol$^{-1}$ for the amino acids studied. If it is assumed that the side chains buried in the interior of the protein have zero configuration entropy, then those entropy values can be considered to approximate the configurational entropy change of transferring a side chain from the interior of the protein to its surface.

Finally, the entropy change gained by the backbone is a function of the steric constraints imposed by the amino acid side chains. $\Delta S_{bb}$ is maximal for glycine and is on the order of 5.5 cal K$^{-1}$ mol$^{-1}$.\textsuperscript{39} The presence of a $\beta$ carbon reduces $\Delta S_{bb}$ by $-2.4$ cal K$^{-1}$ mol$^{-1}$ according to Nemethy et al.\textsuperscript{38} The presence of additional side-chain constituents is expected to have an almost negligible effect on the degrees of freedom of the peptide backbone, except perhaps for valine and isoleucine where the branching at the $\beta$-carbon position might induce additional steric restrictions. Nemethy et al.\textsuperscript{38} estimated that the added steric restrictions in valine and isoleucine may contribute an additional $-2.4$ cal K$^{-1}$ mol$^{-1}$ to $\Delta S_{bb}$. An analysis of the restrictions in backbone entropy by the presence of

disulfide bridges has been presented before. If these values are used to estimate the configurational entropy change for protein unfolding (taking into consideration the amino acid composition and the fact that approximately 50% of the side chains are buried from the solvent), a value extremely close to the experimental value at 112 ° (4.3 cal K⁻¹ mol⁻¹) is obtained.

**Protonation Effects**

Another important contribution to the folding/unfolding energetics is given by the presence of protonatable groups in the protein. Because the pK values of these groups are generally a function of the structural state, changes in pH will affect the stability and cooperative behavior of the protein. Under these conditions, the free energy needs to include the protonation terms, and the folding/unfolding partition function becomes

\[
Q = 1 + \sum_{i=1}^{N-1} \left[ \exp\left(-\frac{\Delta G_i^0}{RT}\right) \prod_{j=1}^{m} \left(1 + K_{ij}a_H\right)(1 + K_{nj}a_H)^{-1} \right]
\]

\[
+ \exp\left(-\frac{\Delta G_N^0}{RT}\right) \prod_{j=1}^{m} \left(1 + K_{Nj}a_H\right)(1 + K_{nj}a_H)^{-1}
\]

which is similar in form to Eq. (3) except that each statistical weight is modified by the terms under the multiplication sign. For each state i, the term under multiplication sign runs over all protonatable groups j and is a function of the protonation constants \(K_{nj}, K_{nj},\) and \(K_{Nj},\) and the hydrogen ion activity \(a_H.\) The population of each state is given by an equation similar to Eq. (4) and the average enthalpy function by Eq. (5a). It must be noted that, in this case, the enthalpy of each state, \(\Delta H_i,\) is given by

\[
\Delta H_i = \Delta H_i^0 + \sum_j \left( F_{ij}\Delta H_{B,ij} - F_{nj}\Delta H_{B,nj} \right)
\]

where \(\Delta H_{B,ij}\) is the protonation enthalpy for site j in state i, and \(F_{ij}\) is its fractional degree of saturation \([F_{ij} = K_{ij}a_H/(1 + K_{ij}a_H)].\) The protonation enthalpies, \(\Delta H_{B,ij},\) are effective enthalpies, in the sense that they are equal to the intrinsic protonation enthalpy of the protein group minus the ionization enthalpy of the buffer system in which the transition is measured. A rigorous analysis of the protonation behavior can be done by performing experiments in buffers characterized by different enthalpies of ionization. In addition, protonation contributions to the folding/unfolding enthalpy can be canceled by the appropriate choice of buffer. It must

---

be noted that the enthalpy contributions arising from protonation are usually small when compared to the total enthalpy of unfolding; they are on the order of around $-1$ kcal mol$^{-1}$ per carboxylic group and about $-7$ kcal mol$^{-1}$ per histidyl group. For comparison, the ionization enthalpies of some common buffers are as follows: $1.2$ kcal mol$^{-1}$ for phosphate, $3.73$ kcal mol$^{-1}$ for MES, $5.32$ kcal mol$^{-1}$ for MOPS, $7.66$ kcal mol$^{-1}$ for Tricine, $11.35$ kcal mol$^{-1}$ for Tris, $0.11$ kcal mol$^{-1}$ for acetate, and $2.74$ kcal mol$^{-1}$ for PIPES (H. Fukada and K. Takahashi, personal communication, 1988).

**Gibbs Free Energy**

The contributions to $\Delta G$ can be written in terms of the $\Delta H$, $\Delta S$, and $\Delta C_p$ changes, structurally parametrized as described above. In addition, other more specific changes including the protonation of ionizable groups, salt bridges, disulfide bonds, or changes in the number of bound ligands need to be accounted individually for each specific case. The contribution of each interaction to the free energy of stabilization can be evaluated from the crystal or solution structure in conjunction with the set of fundamental thermodynamic values ($\Delta H$, $\Delta S$, and $\Delta C_p$) that characterize each type of interaction.

In principle, there are many ways of partitioning the different contributions to the free energy of stabilization of a protein. The parametrization of the energetic parameters discussed above provides an intermediate level description in the sense that the relevant interactions are not partitioned into their most elementary components but rather expressed as composite (effective) quantities containing various elementary contributions. This partitioning results in a relatively small set of effective interactions (polar, apolar, ionic, binding, etc.) that can be related directly to structural parameters as described above.

**Structural Deconvolution of Heat Capacity Function**

In the previous section we discussed the essential elements for a structural parametrization of the folding/unfolding energetics of proteins. The approach presented there can be used to estimate the energetic differences between any two arbitrary states of a protein and, therefore, can be incorporated into the partition function in order to examine the cooperative folding behavior of a protein, the population of partly folded intermediates, or other system parameters.

As mentioned before, a two-state transition is described by three parameters ($\Delta H$, $\Delta S$, and $\Delta C_p$). In general, a transition involving ($N + 1$) states is described by $3N$ parameters, one set of $\Delta H$, $\Delta S$, and $\Delta C_p$ values...
for each additional state. According to the structural parametrization of the energetics, after correction or in the absence of enthalpic contributions from ionization or other ligand binding effects, three structural parameters are required to evaluate $\Delta H$, $\Delta S$, and $\Delta C_p$ for each state: $\Delta ASA_{ap}$, $\Delta ASA_{pol}$, and $\Delta S_{conf}$. Because the number of parameters is the same, it follows that it should be possible to analyze the heat capacity function directly in terms of structural parameters. The analytical method aimed at obtaining structural parameters from the heat capacity function will be called the structural deconvolution of the heat capacity function.

It is easy to show that the changes in solvent accessible polar and apolar surface areas between two protein states can be written as

$$\Delta ASA_{ap} = \frac{\Delta C_p}{a b'} - \frac{\Delta H(60)}{a' b(60)}$$

$$\Delta ASA_{pol} = \frac{\Delta C_p}{a b'} - \frac{\Delta H(60)}{a' b(60)}$$

$$\Delta S_{conf} = \Delta S(T_m) + \Delta C_p \ln(T_m/385.15)$$

In Eqs. (24)–(26) $T_m$ is given in kelvins, $\Delta ASA_{pol}$ and $\Delta ASA_{ap}$ in $\text{Å}^2$, $\Delta H(T_m)$ in cal mol$^{-1}$, and $\Delta C_p$ and $\Delta S_{conf}$ in cal K$^{-1}$ mol$^{-1}$. Figure 3 compares the changes in solvent accessible surface areas obtained from the
Calorimetric data and those calculated directly from the crystallographic structure of the proteins in the thermodynamic database. As seen in the Fig. 3, the agreement is excellent, exhibiting a standard deviation of 4.4% for the polar and 6.8% for the apolar areas (the standard errors being 1.2 and 1.8%, respectively).

The above results indicate that the deconvolution equations can be applied directly in terms of structural parameters and that the excess heat capacity function can be deconvoluted directly to yield $\Delta A_{\text{ASA pol}}$, $\Delta A_{\text{ASA ap}}$, and $\Delta S_{\text{conf}}$. This is illustrated in Fig. 4 for the thermal unfolding of hen egg white lysozyme at two different pH values. In Fig. 4 the experimental excess heat capacity curves have been plotted together with the curves obtained by performing the structural deconvolution of the data. The theoretical curves were generated with the best fitting parameters. The results of the structural deconvolution yield $\Delta A_{\text{ASA pol}} = 4823 \pm 100 \, \text{Å}^2$ and $\Delta A_{\text{ASA ap}} = 6736 \pm 190 \, \text{Å}^2$. For comparison, the changes in accessible surface areas calculated from the crystallographic structure are $\Delta A_{\text{ASA pol}} = 5473 \, \text{Å}^2$ and $\Delta A_{\text{ASA ap}} = 6844 \, \text{Å}^2$. These results indicate that DSC data can be used directly to estimate changes in solvent accessible surface areas. A comparison of these values with those calculated from the crystallographic or solution structure provide a measure of the degree

\[ \begin{align*}
\text{Fig. 4. Structural deconvolution of the excess heat capacity function. The calorimetric data obtained for lysozyme at pH 2 and 2.5 were analyzed directly in terms of changes in polar (} \Delta A_{\text{ASA pol}} \text{) and apolar (} \Delta A_{\text{ASA ap}} \text{) solvent accessible surface areas and } \Delta S^\star \text{. The theoretical curves (solid lines) were generated with the best fitting parameters. The results of the structural deconvolution yield } \Delta A_{\text{ASA pol}} = 4823 \pm 100 \, \text{Å}^2 \text{ and } \Delta A_{\text{ASA ap}} = 6736 \pm 190 \, \text{Å}^2. \text{ For comparison, the changes in accessible surface areas calculated from the crystallographic structure are } \Delta A_{\text{ASA pol}} = 5473 \, \text{Å}^2 \text{ and } \Delta A_{\text{ASA ap}} = 6844 \, \text{Å}^2. \end{align*} \]
of unfolding of a protein after thermal denaturation. Furthermore, if resid-
ual structure is present after thermal denaturation, it is possible to evaluate
the actual persistence of polar and apolar interactions.

Structural Evaluation of Partition Function

The formalism described above allows evaluation of the Gibbs free
energy difference between specific structural states of a protein and the
reference state, which is taken to be the native structure. Because the
total number of possible structural states is astronomical, any realistic
attempt to evaluate the folding/unfolding partition function requires selec-
tion of a subset that is amenable to computation and that can be considered
to have the highest probability of occurrence.

Fortunately, the number of states that ever becomes populated is
relatively small, even under conditions that maximize the population of
intermediates. Furthermore, partly folded states are usually derivatives
of the native structure. It is apparent that the folding/unfolding partition
function can be simplified so that it includes only those states that are
relevant to the folding process. The approach that we have under-
taken involves using the native conformation as a template to gener-
ate partially folded conformations and evaluating the Gibbs free energy
of those conformations according to the rules described above.

The general strategy to approximate the folding/unfolding partition
function from the molecular structure of the folded state involves (1)
generation of the set of most probable partly folded states using the crystal-
lographic structure as a template and (2) evaluation and assignment of the
relative Gibbs free energies of those states. Once this task is accomplished,
the evaluation of the partition function is in principle straightforward. It
has been shown that this approximation to the partition function success-
fully predicts the structural determinants of the folding intermediates of
α-lactalbumin, barnase, staph. nuclease T4 lysozyme, IIIGlc, 434 repressor
and interleukin 1β, proteins for which NMR characterization of the inter-
mediates is available.

Molten Globule State

A third protein state, known as the molten globule state, has attracted
the attention of many investigators owing to the general belief that it
represents a universal folding intermediate. In our own work, we use the
term molten globule or compact denatured state interchangeably to denote

41 E. Freire and D. Xie, Biophys. Chem. in press (1994).
a protein state characterized by (1) significant secondary structure; (2) significant compactness owing to the presence of a sizable hydrophobic core; and (3) a tertiary structure reminiscent of the native fold but without necessarily exhibiting the strict packing constraints of the native state. The structural characteristics of this state have been discussed elsewhere.\textsuperscript{7,42,43}

It has become 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.\textsuperscript{44-47} Also, the exact amount of residual structure is not the same for all molten globules, some of them being almost nativelike and other ones closer to the unfolded state.\textsuperscript{45} This structural diversity underscores the need for an identification of the thermodynamic characteristics of this state.

For several years, the nature and magnitude of the forces that stabilize the molten globule state were a matter of controversy, primarily because direct calorimetric measurements could not be accurately performed in those systems.\textsuperscript{20,48-50} The situation has changed, however, and direct calorimetric measurements of the thermal unfolding of the molten globule of $\alpha$-lactalbumin and cytochrome $c$ have been performed.\textsuperscript{51,52}

Figure 5 shows the temperature dependence of the heat capacity function of $\alpha$-lactalbumin. It is clear that the overall transition proceeds in two stages: a first stage in which the native state denatures cooperatively into a state (molten globule state) with a heat capacity lower than that of the unfolded state and a second stage in which the molten globule gradually unfolds and approaches the heat capacity of the unfolded state. For $\alpha$-lactalbumin, the transition from the folded state to the molten globule state is a highly cooperative process which approaches the behavior of a two-state transition, whereas the transition from the molten globule state to the unfolded state occurs gradually and with a barely noticeable heat effect.

Hierarchical Cooperative Model

It can be assumed that the two-stage character of the thermal unfolding transition is a consequence of the existence of long- and short-range cooperative interactions within the protein molecule. The interactions within the native state of a protein can be used to define an intricate hierarchical network, responsible for the long-range or global cooperative behavior of the molecule. At the most fundamental level are those elements that fold in the absence of other elements. The hydrophobic core of a protein, for example, may fall under this category. Higher order cooperative elements are formed by interactions between lower order elements. For any given protein, the number and extent of the low-order structures will be a function of temperature and solvent conditions. Under certain conditions low-order structures can be more stable than high-order structures and may serve as nucleation centers during the folding process. Conversely, during the unfolding process those same elements will give rise to the appearance of residual structure once the long-range cooperative behavior is disrupted.

From a thermodynamic point of view the classic two-state folding/unfolding behavior is achieved when high-order structures have a higher
stability than low-order structures. In this case, the breakdown of the native structure results in the complete unfolding of the protein. If, on the other hand, a subset of low-order structures have a higher stability than the native state, denaturation will not result in complete unfolding. Under those conditions, denaturation will produce an ensemble of conformations with different degrees of residual structure. The number of conformations in the ensemble is also variable. It could be a single conformation, in which case the overall transition will be a three-state transition, or it could be a very large number of conformations defining an almost continuum of states, in which case the transition will be noncooperative and proceed without a noticeable peak in the heat capacity function. This model is illustrated in Fig. 6. In Fig. 6 the right-hand column represents the entire ensemble of conformations that define the denatured state. At the top of the column is the first state that lacks global cooperativity ($I_0$) and at the bottom the completely unfolded state denoted by U. Depending on the solvent conditions and the temperature at which the transition occurs, the native state will denature to a particular distribution of denatured conformations. In general, as the temperature increases the distribution progressively shifts toward conformations having fewer elements of residual structure until a temperature is reached in which the unfolded state is the predominant species. This model is called the hierarchical cooperative (HC) model. In the case of $\alpha$-lactalbumin, a published structure based analysis of the molten globule suggests that the partly folded states on the right-hand column maintain most of the hydrophobic core defined by the A, B, and C helices.

Statistical Thermodynamic Formalism

The partition function for the HC model is defined in the usual way as

$$Q = 1 + \sum_{i=I_0}^{U} \exp(-\Delta G_i/RT)$$  \hspace{1cm} (27)$$

where $\Delta G_i$ is the Gibbs free energy of state $i$ relative to that of the native state. According to the considerations illustrated in Fig. 6, all the states included inside the summation sign lack global cooperativity. The mathematical implication of this feature of the HC model is that all terms inside the summation can be written as the product of two terms, one of which is the statistical weight of the first state that lacks global cooperativity.
Fig. 6. Schematic representation of the hierarchical cooperative (HC) model for protein folding/unfolding. The left-hand column represents the unique native state conformation. The right-hand column represents the ensemble of denatured states. By definition, all these states lack global cooperativity and are characterized by varying degrees of residual structure. At the top of the column is the first state that lacks global cooperativity ($I_0$). At the bottom of the column is the completely unfolded state (U), which is structureless and fully solvated. The middle of the column comprises denatured states with varying degrees of residual structure. Thermal denaturation can be viewed as a transition from the native state to a particular distribution of denatured molecules. Under certain solvent conditions, the unfolded state is the predominant species, and the transition approaches the classic two-state transition observed for many globular proteins. Under other conditions, the predominant species could be the molten globule state. The molten globule is not necessarily a unique state but a subensemble of states. In this case, a highly cooperative transition will be observed between the native and the molten globule state. On further heating of the protein sample, the center of the distribution will gradually shift toward the unfolded state, giving rise to a noncooperative transformation of the molten globule state into the unfolded state. (Reprinted from Griko et al.51)
If the statistical weight of state \( I_o \) is factorized outside the summation sign, Eq. (27) becomes

\[
Q = 1 + \left[ \exp\left(-\Delta G_{I_o}/RT\right) \right] \sum_{i=I_o}^{U} \exp\left(-\Delta G'_i/RT\right) \tag{28a}
\]

\[
= 1 + \left[ \exp\left(-\Delta G_{I_o}/RT\right) \right] Q_D \tag{28b}
\]

where \( \Delta G_{I_o} \) is the relative Gibbs free energy of the region(s) of the molecule that exhibits long-range cooperativity and \( Q_D \) the internal partition function of the denatured state. If the region(s) that exhibits long-range cooperativity extends to the entire molecule, then \( Q_D = 1 \) and Eq. (28b) reduces to the two-state partition function. The internal partition function for the denatured state can be written as

\[
Q_D = \sum_{i=I_o}^{U} \exp\left(-\Delta G'_i/RT\right) \tag{29a}
\]

\[
Q_D = 1 + \sum_{i=I_1}^{U} \exp\left(-\Delta G'_i/RT\right) \tag{29b}
\]

where the summation runs over all the conformations that define the denatured state. The Gibbs free energies \( \Delta G'_i \) are referenced to the first state in the ensemble of denatured conformations \( (I_o) \) rather than the native state.

As defined by Eq. (29), the number of accessible states having different amounts of residual structure is arbitrary; that is, different proteins may have different numbers of residual structure elements. There are, however, two extreme situations. If the number is equal to one, the total folding/unfolding partition function will reduce to a three-state partition function:

\[
Q = 1 + \exp\left(-\Delta G_{I_o}/RT\right) + \exp\left(-\Delta G_{I_o}/RT\right) \exp\left(-\Delta G'_U/RT\right) \tag{30a}
\]

\[
Q = 1 + \exp\left(-\Delta G_{I_o}/RT\right) + \exp\left(-\Delta G'_U/RT\right) \tag{30b}
\]

In this case, the compact denatured state transforms into the unfolded state in an all-or-none fashion, and depending on the magnitude of the enthalpy difference between the unfolded and compact denatured state, the transformation might give rise to a significant heat absorption peak in the calorimeter. This situation is usually seen in proteins having two structural domains and illustrates the fact that, from a thermodynamic point of view, this situation represents a special case of the more general treatment presented here. The experimental data corresponding to this situation as well as those corresponding to a few discrete states can be
analyzed with the standard deconvolution equations for a multistate transition.\(^1\) In particular, the average excess enthalpy is given by

\[
\langle \Delta H \rangle = \sum P_i \Delta H_i
\]  

(31)

where \(P_i\) is equal to the population of molecules in state \(i\) and is given by

\[
P_i = \exp(-\Delta G_i/RT)/Q
\]  

(32)

The excess heat capacity function obtained by differential scanning calorimetry is simply the temperature derivative of the average excess enthalpy \(\langle \Delta C_p \rangle = \partial \langle \Delta H \rangle / \partial T\).

The second extreme is given by the case in which the denatured state comprises a large number of accessible states. In this case the denatured state defines a continuum of substates separated by very small enthalpy differences. The transformation of the molten globule state into the unfolded state proceeds gradually in a noncooperative fashion and with a minimal excess heat capacity. During this process, the enthalpy of the denatured state increases almost continuously as the distribution of states progressively shifts toward the unfolded state. In this case, the free energy difference between contiguous substates can be approximated by a differential \((\delta G)\), and the partition function becomes

\[
Q = \frac{1}{Q^N} \left[ \exp(-\Delta G_{1o}/RT) \right] \sum_{i=0}^{N} \frac{N!}{i!(N-i)!} \exp(-i\delta G/RT)
\]  

(33a)

\[
= 1 + \left[ \exp(-\Delta G_{1o}/RT) \right] \left[ 1 + \exp(-\delta G/RT) \right]^N
\]  

(33b)

In Eq. (33a) the summation runs over the total number of independent residual structure elements in the denatured state. The intrinsic free energy of a substate with \(i\) residual structural elements is simply \(i\delta G\) since the free energy difference between contiguous substates is assumed to be the same. The combinatorial terms are equal to the number of substates having \(i\) residual structural elements out of a total of \(N\) elements. Because these terms correspond to the binomial expansion, Eq. (33a) can be written in closed form as indicated in Eq. (33b). In this case the average excess enthalpy function is equal to

\[
\langle \Delta H \rangle = \frac{1}{Q} \left[ \exp(-\Delta G_{1o}/RT) \right] \left[ 1 + \exp(-\delta G/RT) \right]^N \Delta H_{1o}
\]  

\[+ N \left[ \exp(-\Delta G_{1o}/RT) \right] \left[ \exp(-\delta G/RT) \right] \left[ 1 + \exp(-\delta G/RT) \right]^{N-1} \delta H \]  

(34)

and, as before, the excess heat capacity function is the temperature derivative of \(\langle \Delta H \rangle\). It must be noted that for \(N = 0\) the HC model [Eqs. (33) and (34)] is equivalent to the two-state model and for \(N = 1\) it corresponds
exactly to the three-state model. For higher \( N \) values, the assumption that the free energy difference between contiguous substates is the same is the most conservative, given that the unfolding of the molten globule state appears to be a very broad process with no features revealing the presence of dissimilar states. The parameter \( N \) is a measure of the cooperativity of the transition from the molten globule to the unfolded state. In fact, from a purely phenomenological point of view, the parameter \( N \) is equal to the calorimetric to van't Hoff enthalpy ratio (\( \Delta H_{cai}/\Delta H_{vH} \)) for the unfolding of the compact denatured state.

**Thermodynamic Properties of Molten Globule of \( \alpha \)-Lactalbumin**

The analysis of the heat capacity function of \( \alpha \)-lactalbumin has revealed several thermodynamic features of the molten globule of this protein. (1) The heat capacity of the molten globule (950 cal K\(^{-1}\) mol\(^{-1}\)) is intermediate between that of the native state and the unfolded state (1800 cal K\(^{-1}\) mol\(^{-1}\)). (2) The molten globule is enthalpically higher than the unfolded state at temperatures below 45°. (3) The transition from the molten globule to the unfolded state is rather noncooperative, suggesting that the molten globule is not a unique state but a collection of states.

The heat capacity of the molten globule state indicates that this state preserves a significant hydrophobic core. On complete unfolding, \( \alpha \)-lactalbumin exposes on the order of 6780 Å\(^2\) of apolar surface and 4750 Å\(^2\) of polar surface to the solvent. These values are consistent with a heat capacity change on the order of 1800 cal K\(^{-1}\) mol\(^{-1}\) as observed experimentally under conditions in which the protein undergoes complete unfolding (e.g., at pH 5.2). The heat capacity difference between the molten globule state and the native state is on the order of 950 cal K\(^{-1}\) mol\(^{-1}\). This figure is consistent with the presence of a significant hydrophobic core, approximately half the size of the one found in the native state.

The most significant observation is that the molten globule state is enthalpically higher than the unfolded state at low temperatures. At high temperatures, however, the unfolded state is enthalpically higher because it has a larger heat capacity, and, therefore, its enthalpy increases at a faster rate with temperature. For \( \alpha \)-lactalbumin the enthalpy of the molten globule and the enthalpy of the unfolded state are equal at approximately 45°. This experimental observation implies that, at low temperatures, the stabilization of the molten globule is necessarily of an entropic origin. More precisely, because the unfolded state has a higher configurational entropy, the main contribution to the stabilization entropy of the molten globule state of \( \alpha \)-lactalbumin must be due to the hydrophobic effect. This observation appears not to be unique to \( \alpha \)-lactalbumin. In fact, experi-
mental studies on the molten globule states of cytochrome c\textsuperscript{13,52} and the heat-labile enterotoxin from \textit{Escherichia coli}\textsuperscript{54} have also revealed that the molten globule state of these proteins has a lower heat capacity than the unfolded state, and that at low temperatures it has a higher enthalpy than the unfolded state. If the molten globule state represents an early folding intermediate, then the initial driving force for folding must be primarily entropic, more explicitly the entropy gain due to solvent release.

The transition from the molten globule state to the unfolded state of \( \alpha \)-lactalbumin is rather noncooperative. In fact, under conditions in which the molten globule state is significantly populated, the best fitting parameters are consistent with a \( N \) value on the order of 2–3. This range of values is significantly larger than the value of 1 expected for a three-state transition. The net effect of this lack of cooperativity is that the transition proceeds with little transition excess heat capacity\textsuperscript{7}, approaching the macroscopic behavior of a transition characterized by no discontinuity in the first derivative of the thermodynamic potential; that is it resembles a higher order phase transition. It must be remembered that the heat capacity function is composed of two terms: the transition excess heat capacity responsible for the characteristic bell-shaped transition peaks associated with protein denaturation and the S-shaped baseline shift due to the existence of a positive \( \Delta C_p \). The latter term is an ensemble average directly proportional to the population of molecules in the denatured and unfolded states. Owing to the absence of a significant excess heat capacity, the measured heat capacity does not exhibit a clear peak and seems to monotonically approach the heat capacity of the unfolded state as the temperature increases (in some cases, the excess heat capacity function may look like an S-shaped curve). Throughout this gradual increase in heat capacity the enthalpy of the system increases as the distribution of molecules progressively shifts toward the unfolded state. During this process, it is possible for some system parameters to display transitions that occur in a narrower temperature range\textsuperscript{55,56}. This is reminiscent of order–disorder transitions or percolation phenomena in which a totally noncooperative transformation gives rise to extremely sharp phase transitions in certain system parameters like the electrical resistivity in alloys\textsuperscript{57}. In proteins,

\textsuperscript{53} K. P. Murphy, A. D. Robertson, N. D. Meadow, S. Roseman, and E. Freire, manuscript in preparation (1994).
\textsuperscript{54} D. Xie, V. Bakhuni, and E. Freire, manuscript in preparation (1994).
as the degree of unfolding of the denatured state increases, the molecule undergoes a transition from a collapsed conformation to an extended one. This change of state is not necessarily a linear function of the degree of unfolding and might occur within a small domain of the excess enthalpy and, as such, be accompanied by a relatively small enthalpy change.

In general, the cooperativity of the transition from the molten globule state to the unfolded state is not necessarily similar for all proteins. Some proteins can be expected to exhibit rather cooperative transitions, as in the case of cytochrome c or equine lysozyme, and some other proteins very diffuse transitions, as in the case of apomyoglobin.

Concluding Remarks

In this chapter, advances in the structural thermodynamic analysis of protein stability have been discussed. This analysis involves the linkage of thermodynamic and structural parameters in a quantitative way using the partition function formalism. The validity of this approach can be seen in its ability to predict experimental system properties.

Acknowledgments

Supported by grants from the National Institutes of Health (RR-04328, GM-37911, and NS-24520) and the National Science Foundation (MCB-9118687).


By Martin Straume

Introduction

Elucidation of the detailed energetics of protein folding-unfolding equilibria together with knowledge of specific structural information provides a powerful means for gaining a productive understanding of the principles