Protein Stability Curves

WAYNE J. BECKTEL and JOHN A. SCHELLMAN, *Institute of Molecular Biology and Department of Chemistry, University of Oregon, Eugene, Oregon 97402*

Synopsis

The stability curve of a protein is defined as the plot of the free energy of unfolding as a function of temperature. For most proteins the change in heat capacity on denaturation, or unfolding, is large but approximately constant. When unfolding is s two-state process, most of the salient features of the stability curves of proteins can be derived from this fact. A number of relations are obtained, including the special features of low-temperature denaturation, the properties of the maximum in stability, and the interrelationships of the characteristic temperatures of the protein. The paper closes with a formula that permits one to calculate small changes in stabilization free energy from changes in the melting temperature of the protein.

INTRODUCTION

A knowledge of the origins of the stability of proteins is essential to the understanding of their structure and function. The stability must be great enough for the protein to find and maintain its native conformation relative to other conformations, but not so great that conformational changes or adjustments, considered an integral part of many protein functions, are precluded. There are a number of quantitative measures of stability. In the earlier literature protein stability was often tested by subjecting a protein to high temperatures in open vessels for varying periods of time, and testing for insolubility or the recovery of activity. We now know that this type of procedure depends on irreversible processes, both chemical and physical, and therefore has kinetic as well as equilibrium aspects. Though this measure of stability is of great practical importance, it is not the kind of stability that will be of interest in this paper. Another measure of stability is the depth of the energy minimum calculated by energy minimization programs. The drawbacks of this method are that the potential functions are not entirely accurate and that the solvent is usually ignored. In this paper stability will be given a thermodynamic definition. This is the difference between the partial molar free energy of the macromolecule in its unfolded or denatured state (u) and its native folded state (f).

$$\Delta\overline{G} = \overline{G}_u - \overline{G}_f \tag{1}$$

The notation used in Eq. (1) and below is meant to distinguish between molar quantities and system-extensive variables such as G, H, etc., rather than between partial molar quantities and apparent molar quantities. If the unfold-

Biopolymers, Vol. 26, 1859–1877 (1987)
© 1987 John Wiley & Sons, Inc.
CCC 0006-3525/87/111859-19$04.00

ing reaction can be represented as a two-state process, this free-energy change is experimentally measurable.[1-5] Some authors list the free energy of the folded state relative to the unfolded state so that the stabilization free energies are negative stable structures. It should be noted that the unfolded and folded states both contribute explicitly to this definition of protein stability. There is a temptation, especially with proteins of known crystal structure, to relate changes in stability exclusively to features of the native structure of the molecule. This mode of thought must be avoided because it is likely that a large component of the free energy of stabilization as defined above stems from the increased solvation of the unfolded chain relative to the folded one. Though the inclusion of both states leads to uncertainties in the interpretation of some of the experimental results, it is clearly the free energy of the folded state relative to the unfolded state generated at the ribosome that drives the formation of secondary and higher order structure of proteins. As a result, the thermodynamic definition, apart from the differences between in vivo and laboratory conditions, is directly relevant to the biological process of protein folding.

The stabilization free energy depends on the usual thermodynamic variables of temperature, pressure, and composition. Usually, the important composition variables for protein stability are pH, ionic strength, the concentration of denaturants, and the concentration of other substances that bind preferentially to the native or unfolded states. The effect of pressure will not be discussed in this paper, but we shall broaden the range of variables to include alterations in primary structure caused by mutations (substitution of one amino acid for another in the sequence) or chemical modification.

DEPENDENCE ON TEMPERATURE AND STABILITY CURVES

The earliest thermodynamic models for protein denaturation made use of the standard first approximation of thermochemistry by assuming that $\Delta\bar{C}_p = 0$ for the reaction, so that the enthalpy and entropy of unfolding are independent of temperature. With the advent of the hydrophobic effect and its singular temperature dependence,[6,7] it became clear that this was a very poor representation of the thermodynamics of unfolding (see Fig. 1). In an important series of investigations, Brandts[8] determined free energies of denaturation at different temperatures and pH for chymotrypsinogen. The temperature variation of the enthalpy and entropy was found to be quite large. The nature and accuracy of this temperature variation were examined by Shiao, Lumry, and Fahey.[9] Later, in a series of studies of unprecedented precision, Privalov and Khechinashvii established that $\Delta\bar{C}_p$ of protein unfolding is in fact quite large and positive, and that within experimental error it can be taken as a constant for a given protein.[10] This conclusion has been subsequently supported by many other investigations on many proteins. Constant $\Delta\bar{C}_p$ is the standard second approximation of thermochemistry. A great deal can be deduced about the stability properties of proteins from just three postulates: (1) the denaturation reaction can be regarded as a two-state process so that $\Delta\bar{G}$, $\Delta\bar{S}$, and $\Delta\bar{H}$ are experimentally defined; (2) the protein is stable at some temperature, so that $\Delta\bar{G}$ is positive over some range of

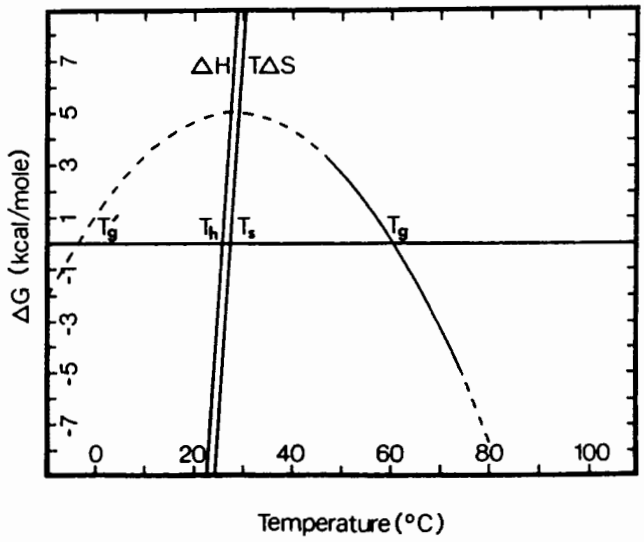

Fig. 1. A stability curve for a protein. The curve is derived from the temperature variation of the enthalpy and entropy of unfolding. The solid line represents that portion of the curve used in a van't Hoff analysis of the transition. The dashed line is extrapolated. An explanation of the temperatures in the figure appears in the text.

temperature; and (3) $\Delta\bar{C}_p$ = constant > 0. With the second postulate we have

$$\Delta\bar{H} = \Delta\bar{H}^0 + \Delta\bar{C}_p(T - T^0) \tag{2}$$

$$\Delta\bar{S} = \Delta\bar{S}^0 + \Delta\bar{C}_p\ln(T/T^0) \tag{3}$$

$$\Delta\bar{G} = \Delta\bar{H}^0 - T\Delta\bar{S}^0 + \Delta\bar{C}_p\left[T - T^0 - T\ln(T/T^0)\right] \tag{4}$$

where T^0 is any reference temperature and $\Delta\bar{S}^0$ and $\Delta\bar{H}^0$ are the changes in partial molal entropy and enthalpy at that temperature, respectively. There are three parameters in these equations ($\Delta\bar{C}_p$, $\Delta\bar{S}^0$, and $\Delta\bar{H}^0$) and the determination of these is sufficient to establish the course of the enthalpy, entropy, and free energy over the temperature range for which the third postulate is valid.

A drawing of the entropy, enthalpy, and free-energy functions, taken from an experimental example, is shown in Fig. 1. We define the curve of $\Delta\bar{G}$ vs T as the stability curve of a protein. The relative stabilities of two systems (either two different proteins, or the same protein under different experimental conditions) are then compared by way of their stability curves. The temperature dependence is thus an inherent part of the comparison. The approach of our laboratory is to describe the effects of pH, reagents, mutations, etc., in terms of the effect that these variables have on the position and shape of the stability curve.

ELEMENTARY PROPERTIES OF THE STABILITY CURVE

The stability curve has some very general features. The curvature is given by $\partial^2\Delta\bar{G}/\partial T^2 = -\Delta\bar{C}_p/T$, which by postulate is negative at all temperatures. The slope of the curve is given by $\partial\Delta\bar{G}/\partial T = -\Delta\bar{S}$. We define the temperature at which the stability is maximum as T_s, which is to be interpreted as the

temperature at which $\Delta\bar{S} = 0$. If T_s is used as the standard temperature, then the entropy may be written as

$$\Delta\bar{S} = \Delta\bar{C}_p \ln(T/T_s) \tag{5}$$

The slope of the stability curve is given by $-\Delta\bar{S}$ and is always negative for $T > T_s$. Therefore the curve has only one extremum, which is a maximum. In our notation the free-energy change at the maximum is written as $\Delta\bar{G}_s$, indicating that the free energy is evaluated at T_s. Throughout this paper subscripts s, h, and g will refer to functions evaluated at temperatures where $\Delta\bar{S}$, $\Delta\bar{H}$, and $\Delta\bar{G}$, respectively, are zero. By postulate the native protein is stable in some range of temperature such that $\Delta\bar{G}$ is positive in that region, and the stability curve crosses the abscissa at two points. These points constitute the high temperature and putative low temperature melting points of the protein. The low-temperature melting point is often extrapolated to be below the freezing point of water. We note also that at T_s, $\Delta\bar{G}_s = \Delta\bar{H}_s$, i.e., that transformation is purely enthalpic.

By definition the melting temperatures are given as the points where $\Delta\bar{G} = 0$, and are denoted by T_g. A prime will be used for the lower inverse melting temperature when a distinction is necessary. Since $\Delta\bar{G} = 0$ at T_g by definition we have

$$T_g = \Delta\bar{H}_g/\Delta\bar{S}_g \tag{6}$$

T_g is usually measured as the midpoint of a calorimetric or spectroscopically measured transition. $\Delta\bar{H}_g$ is obtained as a direct thermal measurement with a calorimeter or as the slope of an $R\ln K$ vs $1/T$ plot in the van't Hoff Analysis. $\Delta\bar{S}_g$ may then be determined from Eq. (6). $\Delta\bar{C}_p$ can be measured directly with calorimetry but normally a procedure of Privalov is used, which involves measuring $\Delta\bar{H}$ as a function of pH.[10] The procedure had been used earlier by Shiao et al. with spectroscopic techniques.[9] The basis for this method will be discussed in the section on pH.

A third characteristic temperature that is sometimes useful is T_h, the temperature at which $\Delta\bar{H} = 0$. We can get relations between these temperatures as follows: Equations (2) and (3) are valid for any pair of temperatures. Since information is usually available first at T_g, it is normally the first reference temperature used in working up the data. With $\Delta\bar{H}_g$ and $\Delta\bar{S}_g$ known, one can calculate T_s and T_h from the formulas

$$T_g - T_h = \Delta\bar{H}_g/\Delta\bar{C}_p \tag{7}$$

$$\ln(T_g/T_s) = \Delta\bar{S}_g/\Delta\bar{C}_p \tag{8}$$

The expressions for the entropy and enthalpy are especially simple when T_s and T_h, respectively, are used for reference temperatures. This has been given as Eq. (5) above for the entropy, and for the enthalpy we have

$$\Delta\bar{H} = \Delta\bar{C}_p(T - T_h) \tag{9}$$

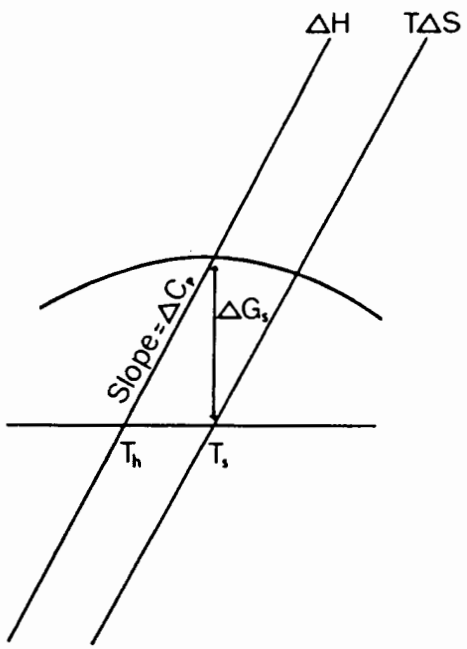

Fig. 2. The triangular relationship between T_h, T_s, and the maximum free energy of stabilization.

Evaluating Eq. (9) at T_s gives a simple relation between T_s and T_h:

$$T_s - T_h = \Delta\bar{H}_s/\Delta\bar{C}_p = \Delta\bar{G}_s/\Delta\bar{C}_p \tag{10}$$

This establishes a simple triangular relationship between $T_s - T_h$ and the maximum stability of the protein $\Delta\bar{G}_s$ (see Fig. 2).

$\Delta\bar{C}_p$ for the unfolding of proteins is quite large and usually falls in the range of 12–18 cal/deg per residue. This leads to an extraordinarily steep dependence of both the entropy and enthalpy of unfolding on temperature (Fig. 1). On the other hand, the maximum of stability of proteins tends to be small, on the order of 50–100 cal/residue. As a result, T_s and T_h tend to be within a few degrees of one another [see Eq. (10) and Fig. 1].

TEMPERATURE VARIATION OF $\Delta\bar{C}_p$

Using the relationships given above it is a simple matter to calculate the cardinal quantities of the stability curve (the two T_gs, T_s, T_h, and the maximum free energy $\Delta\bar{G}_s$) from the standard quantities observed from experiments (T_g, $\Delta\bar{H}_g$, and $\Delta\bar{C}_p$). As shown in Fig. 1, however, this represents a significant extrapolation from the data near T_g. The accuracy of the extrapolation depends, of course, on how well the transition has been measured, as shown in the next section. The validity of the extrapolation, however, is more greatly influenced by how well our three postulates represent the true behavior of the protein over the extended temperature range. In particular, is $\Delta\bar{C}_p$ actually a constant?

From very general principles, this assumption cannot be correct over an extended range of temperatures. The third law of thermodynamics requires both native and denatured heat capacities to vanish at 0° K. In addition, any statistical mechanical description of heat capacities of the unfolded and folded

states of proteins would require incredible coincidences in molecular parameters to maintain a constant difference in heat capacity between two such diverse states of a molecule. The relevant question is, however, How much $\Delta\bar{C}_p$ might vary for proteins in aqueous solutions?

We approach the temperature variation of $\Delta\bar{C}_p$ in terms of both the accuracy and precision of its determination. The question has been previously addressed in van't Hoff[9] and calorimetric[5] measurements. The errors in $\Delta\bar{C}_p$ determinations for chymotrypsinogen[9] were deemed so large as to exclude van't Hoff measurements. Our own determinations of $\Delta\bar{H}(T)$ for several T4 lysozymes indicate errors of 5–10% in the slope of $\Delta\bar{H}_g$ vs T_g. (W. Becktel, D. Muchmore, and W. Baase, in preparation). Errors of this magnitude are reported by Privalov for calorimetric determinations, and we assume the general accuracy of $\Delta\bar{C}_p$ measurements to be of this order of magnitude. The range of T_g available by the variation of pH is seldom greater than 50°C.[5] For a protein with $\Delta\bar{C}_p$ equal to 2 ± 0.2 kcal/deg mole, a systematic temperature dependence of up to 4 cal/deg^2 mole would, therefore, not be observed. As shown below, this is likely an upper bound for the temperature dependence of $\Delta\bar{C}_p$. A better estimate may be obtained by considering the precision of a calorimetric determination.

Calorimetric methods of determining $\Delta\bar{C}_p$ rely upon the precision with which the difference between the specific heats of the native and denatured states may be obtained. For one of the most precise scanning microcalorimeters, the DAS-1M, this is approximately 20 microcal/deg g. This estimate is based upon the 1 mL volume and 20 microcal/deg mL precision in the difference in heat capacity between the sample and reference cells.[10] For a 20 kilodalton protein at a concentration of 5×10^{-3} g/mL, having a $\Delta\bar{C}_p$ of 2 kcal/deg mole, the minimum error in $\Delta\bar{C}_p$ is 4%. This leads, in turn, to a potential systematic variation with temperature of 2 cal/deg^2 mole for $\Delta\bar{C}_p$. Comparing this to the previous estimate for van't Hoff measurement, we conclude that the potential temperature variation of $\Delta\bar{C}_p$ is 5–10%. The question is: How much of an effect would such a change in $\Delta\bar{C}_p$ have on stability curves?

The next standard approximation in thermochemistry is the assumption that $\Delta\bar{C}_p$ varies linearly with temperature. The effect on $\Delta\bar{H}$ is to convert it from a linear to a parabolic function of temperature. Our expression for $\Delta\bar{H}$, $\Delta\bar{S}$, and $\Delta\bar{G}$ then assume the following forms:

$$\Delta\bar{H} = \Delta\bar{H}_{\text{linear}} - \beta/2\left(T_g^2 - T^2\right) \tag{11a}$$

$$\Delta\bar{S} = \Delta\bar{S}_{\text{linear}} - \bar{\beta}\left(T_g - T\right) \tag{11b}$$

$$\Delta\bar{G} = \Delta\bar{G}_{\text{linear}} - \bar{\beta}/2\left(T_g - T\right)^2 \tag{11c}$$

In these equations the term *linear* refers to the temperature dependence of the enthalpy in Eq. (9). From these equations it is apparent that if $\bar{\beta}$ is positive, $\Delta\bar{G}$ is reduced at temperatures both above and below the reference temperature, and conversely for negative values of $\bar{\beta}$. An example of the effect of a 10% linear variation of $\Delta\bar{C}_p$ is shown in Fig. 3. Note that when $\bar{\beta}$ is

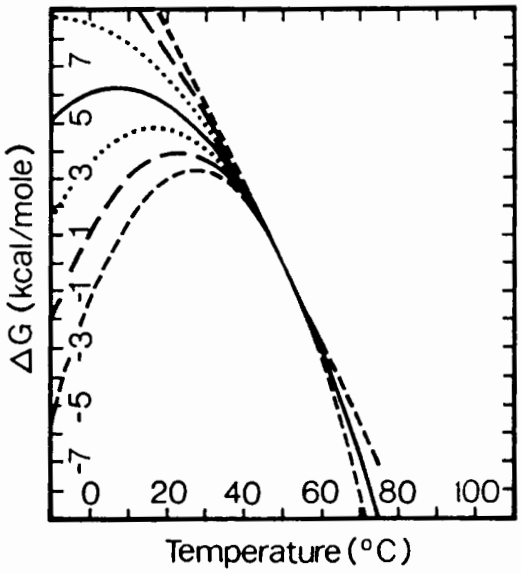

Fig. 3. The effect of linear temperature variation of $\Delta\bar{C}_p$. The solid line represents the derived variation of $\Delta\bar{G}$ with temperature. The enthalpy used to construct this curve was 91.5 kcal/mole, the melting temperature 50°C, and $\Delta\bar{C}_p$ of 2 kcal/deg mole. Positive and negative values for the temperature dependence of $\Delta\bar{C}_p$ were used to construct the remaining lines. These were 6 (short dash), 4 (long dash), and 2 (dots) cal/deg^2 mole. Negative values yield curves that lie below the solid line.

negative and $\Delta\bar{C}_p$ decreases with temperature, the extrapolated curves are raised and T_g' becomes even more inaccessible to observation than the predicted curve for constant $\Delta\bar{C}_p$. On the other hand, when $\bar{\beta}$ is positive the curve is lowered. If $\Delta\bar{C}_p$ were to increase 10%, from 25–75°C, then low-temperature denaturation would be expected to occur. This is not commonly observed in the absence of denaturants, and it would seem that a temperature variation of this magnitude and sign is unlikely. Without more accurate measurements of $\Delta\bar{C}_p$ it is not as easy to rule out instances where it decreases with temperature.

EXTRAPOLATED STABILITY CURVES

Although experiments vary in accuracy, it is usually only possible to measure equilibrium constants of unfolding in the range of 0.1–10. This implies that, for temperatures in the neighborhood of 50°C, stabilities are normally only empirically determined for $\Delta\bar{G}$ in the range of ± 1.5 kcal/mole. For this reason many of the special features of the stability curve, such as the stability maximum, the low-temperature transition, and the change in sign of $\Delta\bar{S}$ and $\Delta\bar{H}$, are not directly observed and must be justified. This is true for calorimetric as well as spectroscopic methods since, in most instances, the stability maximum and other parts of the stability curve lie at temperatures outside the range of those where melting takes place. The main assumptions in the extrapolation of $\Delta\bar{G}(T)$ from any of these techniques is that the transition is two state in nature, that $\Delta\bar{C}_p$ is constant, and that $\Delta\bar{H}(T_g)$ and $\Delta\bar{C}_p$ have been accurately determined. The systematic errors in extrapolation caused by the variation of $\Delta\bar{C}_p$ with temperature have been discussed above and errors inherent in calorimetric measurements have been detailed previously.[10] A number of studies, including our own, make use of spectroscopic techniques and van't Hoff plots in measuring protein stability. The

difficulties associated with such measurements have been extensively discussed.[8,9] Briefly, because of the temperature dependence of $\Delta\bar{H}_g$, plots of $\ln K$ vs $1/T$ would be expected to be nonlinear over the range of temperatures at which proteins melt. The usual solution to this problem is to limit the equilibrium constants used to those only a few degrees on either side of T_g. The enthalpies and melting temperatures derived from such an analysis are approximately correct. Our own approach, where the transition is two state in nature and the data of sufficient accuracy, is to carry out this limited analysis but also to carry out nonlinear fits of $\ln K$ vs $1/T$ in which the temperature variation of $\Delta\bar{H}$ is explicitly allowed. We observe that the melting temperatures and enthalpies derived by both methods differ by less than their respective experimental errors.

Error analysis on all determined parameters is routinely obtained in our laboratory. Using this information, we can assume that errors fall within one standard deviation of each of the parameters of a stability curve, and obtain a zone of extrapolation, i.e., a region in which the stability curve is likely to fall. From Eqs. (2)–(4), the error in $\Delta\bar{G}_s$ has the following form:

$$\sigma_g^2 = \sigma_h^2 + \sigma_c^2 A^2 + \sigma_t^2 B^2$$

$$A = (T_s - T_g) - T_s \ln(T_s/T_g) \tag{12}$$

$$B = -\Delta\bar{C}_p + \Delta\bar{H}T_s/T_g^2 + \Delta\bar{C}_p T_s/T_g$$

This equation arises from a routine propagation of errors analysis. The subscripts to the standard deviations in Eq. (12) refer to the free energy (g), the enthalpy (h), $\Delta\bar{C}_p$ (c), and the melting temperature (t). Expressions for the errors in T_s and T_h may be obtained in the same manner. A typical stability curve with a zone of extrapolation is shown in Fig. 4. For the experiment used in the construction of this figure, these values turn out to be

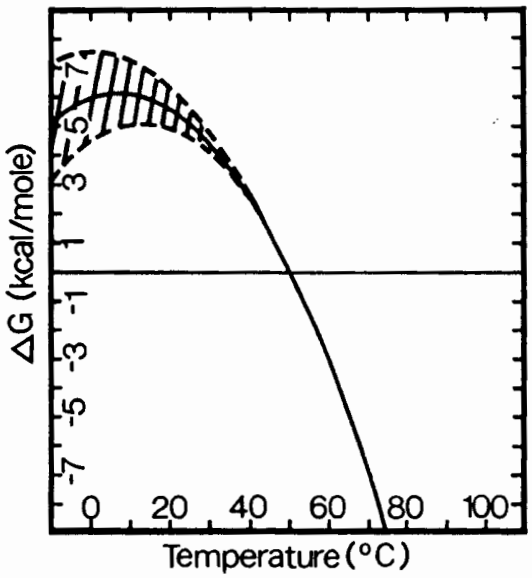

Fig. 4. The effect of errors in T_g, $\Delta\bar{C}_p$, and $\Delta\bar{H}_g$ on a stability curve. The solid curve had the same parameters as Fig. 3. The upper dashed curve has an enthalpy 5 kcal/mole greater and $\Delta\bar{C}_p$ 0.2 kcal/deg mole less than the solid line. In the lower dashed curve the enthalpy is 5 kcal/mole less and $\Delta\bar{C}_p$ 0.2 kcal/deg mole greater than for the solid line.

$\Delta\bar{H}_g = 91.5 \pm 5.0$ kcal/mole, $\Delta\bar{G}_s = 6.9 \pm 0.9$ kcal/mole, $T_g = 50.0 \pm 0.5$°C, $T_s = 7.3 \pm 3.6$°C, and $T_h = 4.3 \pm 2.6$°C.

While the above considerations provide us with a quantitative measure of the accuracy of stabilities obtained by extrapolation, we have in addition good reasons for believing in the qualitative aspects of the stability curve. In the first place plots of $\Delta\bar{H}$ and $\Delta\bar{S}$ as a function of temperature are so steep and dramatic (Fig. 1) over the range where experimental results are obtained that they can fail to change sign for most proteins only if there is a drastic change in the thermodynamic properties of the proteins somewhere between the melting point of the protein and 0°C. For this there is no evidence. In addition to the flattening of the stability curve at low temperature, evidence of a maximum in stability has been observed by all workers in the field and the maximum has actually been observed in a number of instances. As examples we cite the work of Brandts,[11] Pace and Tanford,[12] and Nojima et al.[13]

Finally, in a few special instances the low-temperature transition has actually been observed. The lowering of the stability curve by some external agent has usually been required for these cases. Christensen found a low-temperature melting for β-lactoglobulin in urea solutions.[14] This was confirmed for the same protein by Pace and Tanford[12] in guanidinium chloride solutions. Brandts induced a low-temperature transition in ribonuclease by applying external pressure.[15] Nojima et al.[13] and Becktel et al.[16] have studied the low-temperature melting of yeast phosphoglycerate kinase and phage T4 lysozyme, respectively, in solutions of guanidinium chloride. In the latter cases, complete stability curves, including both transitions and the maximum, have been determined. In addition, Cho et al.[17] and Privalov et al.[18] have observed the low-temperature melting of myoglobin at low pH. It might be maintained in these instances that the additions of reagents such as acid, urea, or guanidinium chloride, or an increase in pressure, actually produce the low-temperature melting by an independent mechanism. This is unlikely in our view. We believe the effect of external agents such as those mentioned is to lower the maximum in the stability curve so that the low-temperature transition and the maximum become accessible to experiment, but without introducing qualitative new features in the shape of the curve. This point can be cleared up by a quantitative study of the effect of external agent on the stability as a function of temperature. Such a study has been performed by B. L. Chen in our laboratory for β-lactoglobulin, with results in agreement with this hypothesis. To our knowledge a low-temperature unfolding has not yet been found for a monomeric protein at neutral pH in the presence of noninteracting salts and buffers, although the low-temperature melting of yeast phosphoglycerate kinase occurs at a guanidinium chloride concentration of only $0.5M$.[13]

AN ALTERNATIVE MEASURE OF STABILITY

Because the free-energy function of a system is a minimum at equilibrium under conditions of constant temperature and pressure, it is understandable that the relative stabilities of two states of the system (or a component) are evaluated in terms of the difference in free energy (chemical potential)

between them. This is the normal procedure in comparing the native and denatured states of proteins. Since the relative free energies of two states are very often measured via the equilibrium constant for their interconversion with the standard thermodynamic formula, $\Delta\bar{G} = -RT\ln K$, it would not be unreasonable to make use of $R\ln K$ itself as a measure of the relative stabilities of two states. In many cases, this turns out to be more intuitive than definitions based on $\Delta\bar{G}$.

This may be seen by way of some elementary examples. Consider the equilibrium reaction $A = B$; $K = [B]/[A]$. The higher the probability of B relative to A, the higher its relative population (concentration) and the higher K and $R\ln K$. If instead we are dealing with a phase equilibrium, say the solubility of a pure substance, then the higher the probability, the higher the solubility. For a pure substance in equilibrium with a liquid phase, the equilibrium constant is the solubility itself and $R\ln K$ is again a good measure of solution tendency or of the stability of the substance in the dissolved state.

The difference between $\Delta\bar{G}$ and $R\ln K$ as a measure of stability is more than a matter of semantics. $\Delta\bar{G}$ has its maximum (or minimum by the alternative definition mentioned in the introduction) where $\Delta\bar{S} = 0$, i.e., at T_s. $R\ln K$, on the other hand, has its maximum (or minimum) value where $\Delta\bar{H} = 0$, i.e., T_h. As we have seen, these two temperatures are separated by $\Delta\bar{G}_s/\Delta\bar{C}_p$ for a model with a constant $\Delta\bar{C}_p$. For protein stability the difference is only a few degrees. Also, $R\ln K$ and $\Delta\bar{G}$ vanish at the same points (T_g and T_g') so that there are relatively minor differences between the respective stability curves.

As a counterexample we may take the recent model of hydrophobic interactions proposed by Baldwin.[19] The model is based on calorimetric data of Gill and his co-workers[20,21] on solutions of hydrocarbons in water, and an observation by Sturtevant[22] that $\Delta\bar{S}(298)/\Delta\bar{C}_p$ is essentially a constant for a number of hydrophobic substances when they are dissolved in water. Baldwin

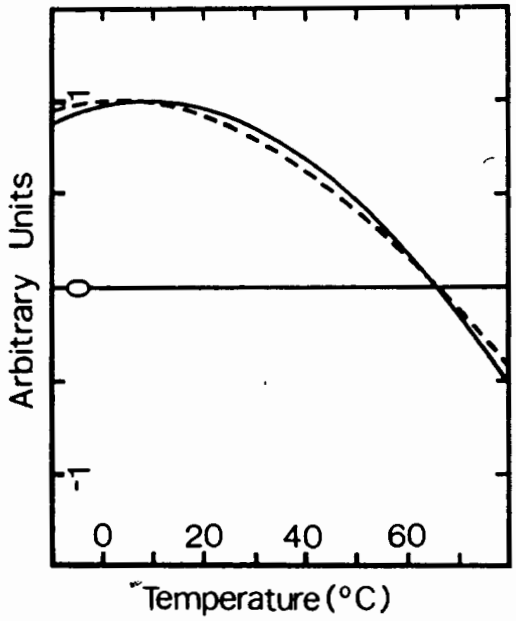

Fig. 5. A comparison of two different measures of stability. The solid curve represents the variation of $\Delta\bar{G}$ and the dashed $-\Delta\bar{G}/T$. Both curves have been scaled so that their respective maxima equal unity. The parameters used to construct these curves were $T_h = 3.4°C$, $T_s = 8.8°C$, $T_g = 65°C$, and $\Delta\bar{G}_s = 10.9$ kcal/mole.

used Eq. (5) to calculate a common T_s for hydrophobic systems, which turned out to be 113°C. This is the temperature of maximum free energy of solution. On the other hand, the solubility of all the hydrocarbons goes through a minimum near 20°C. This is T_h, the temperature at which $-R\ln K$ is a maximum, where K is the solubility. The difference between T_s and T_h is close to 100°C in this instance.

The function $R\ln K$ is equivalent to $-\Delta\overline{G}/T$ and corresponds to one of the free-energy functions proposed by Massieu before the Gibbs and Helmholtz functions were invented. As a criterion of stability $-\Delta\overline{G}/T$ or $\Delta\overline{G}/T$ is more intuitive than $\Delta\overline{G}$ itself. The probability of finding a protein molecule in its unfolded state is least at T_h, not T_s; the solubility of hydrocarbons is least at 20°C, not at 113°C.

We nevertheless continue to use $\Delta\overline{G}$ as a measure of protein stability, but mainly because T_h and T_s are so close to one another for protein denaturation that the representations are essentially equivalent. Figure 5 shows curves of $\Delta\overline{G}$ and $\Delta\overline{G}/T$ that have been scaled to the same amplitude.

VARIATION WITH pH IN THE ACID REGION

A number of years ago Privalov and his co-workers studied the effect of pH on the melting temperature and enthalpy of denaturation for a number of proteins (See Privalov's review, Ref. 5). It was noted that plots of the enthalpy vs the corresponding melting temperature were linear for all proteins studied. It was concluded that in these experiments $\Delta\overline{H}_g$ is a direct function only of the temperature:

$$\Delta\overline{H}_g(T_g) = \text{const} + \Delta\overline{C}_p T_g \tag{13}$$

Mathematically this means that neither $\Delta\overline{H}$ nor $\Delta\overline{C}_p$ in Eq. (2) is a function of pH. The apparent dependence of $\Delta\overline{H}$ on pH results from the fact that pH changes T_g, the temperature at which $\Delta\overline{H}$ is evaluated. The linear relationship between $\Delta\overline{H}_g$ and T_g, when pH is varied, has since been confirmed for many proteins in many laboratories, including our own. This relationship is of great importance for studies in which spectroscopic rather than calorimetric methods are used. The van't Hoff analysis of melting curves is not sufficiently accurate to determine $\Delta\overline{C}_p$ directly. One instead varies the pH to get $\Delta\overline{H}$ at a number of melting temperatures. The slope of the straight line of ΔH_g vs T_g is then used to evaluate $\Delta\overline{C}_p$.

The state of ionization of a protein is changed by both denaturation and changes in pH. The experimental results for $\Delta\overline{H}$ and $\Delta\overline{C}_p$ could depend significantly on the ionic processes that occur. Three factors that should be considered are (1) the change in ionization of the protein groups; (2) the titration of buffer by protons released or taken up by the protein; and (3) the expansion of the denatured protein, which is a function of pH (protein charge) and changes in the electrostatic energy. We can write

$$\Delta\overline{H}_{\text{exper}} = \Delta\overline{H}_{\text{conf}} + \Delta\overline{H}_{\text{pi}} + \Delta\overline{H}_{\text{buff}} + \Delta\overline{H}_{\text{expan}} \tag{14}$$

The subscripts in Eq. (14) stand for the experimental (exper), conformational (conf), protein ionization (pi), buffer titration (buff), and protein expansion

(expan) components of the enthalpy. Pfeil and Privalov,[23,24] and Privalov et al.,[18] have addressed the first and second contributions experimentally, and have shown that it is possible to select buffers so that the enthalpy changes of the protein ionization are effectively canceled by those of the buffer. The experimental enthalpy can then be equated to the conformational enthalpy provided the enthalpy of expansion can be ignored. The latter is probably a good approximation since the enthalpy of charge interactions in water are very small compared to the free energy of interaction.[25]

When a protein unfolds in acid pH there is usually an uptake of protons, which can be determined by the thermodynamic formula

$$\left(\partial 1/T_g\right)/\partial \,\mathrm{pH} = \left(2.3R/\Delta \overline{H}_g\right)\Delta\nu \tag{15}$$

This can occur either because ionizable groups are buried and take up protons when they are exposed in the denatured state (for example, histidine residues at pHs below their pKs[18] or because expansion of the protein changes the effective pKs of the groups. In T4 lysozyme, for example, only the carboxyl groups of glutamic acid and aspartic acid are untitrated in acid solution. The effective pKs of the carboxyls are lowered because of the electrostatic repulsion of the rest of the groups in the protein, which has a high positive charge.[26] When the protein is denatured and expands, the electrostatic repulsion is diminished and 2 to 3 protons are taken up. For the case of myoglobin Privalov et al. have compensated the large enthalpy of ionization of histidines by using piperazine:HCl as a buffer, which has almost the same enthalpy. When the principal protein ionizations are carboxylate groups, acetate buffer provides essentially perfect compensation. The enthalpy effects of carboxylate ionization are at any rate very small.

The effect of pH on $\Delta\overline{G}$ itself was derived by Hermans and Scheraga[27] and is given by the relationship[28]

$$\Delta\overline{G} = \Delta\overline{G}^{\circ} - RT\ln\left(\Sigma^{u}/\Sigma^{f}\right) \tag{16}$$

where $\Delta\overline{G}^{\circ}$ is the free energy of unfolding with no change in ionization. Σ^{f} and Σ^{u} are the binding polynomials of the native and denatured forms of the protein, which can be measured as a function of pH by means of titration curves.

When $\Delta\overline{H}$ shows no direct dependence on pH, T_h is a constant and is therefore a convenient reference temperature for comparing protein stabilities. This also means that the interpretation of Eq. (2) or (9) can be expanded to describe the enthalpy of unfolding of a protein as a function of temperature over a wide range of pHs. We note that, since $\Delta\overline{G}_s$ is purely enthalpic, it must slide up and down the pH-independent straight line of Eq. (2) as the pH is changed. Model calculations illustrating these results are presented in Fig. 6. In the next section we show how T_g is affected by pH, and how changes in T_g can be used to give approximate values of the perturbations in stabilization free energy produced by changes in pH. The relations given in this section are useful in providing a general view of the effect of pH on stability, but the accurate way of getting this information is to do careful stability curves as a function of pH.

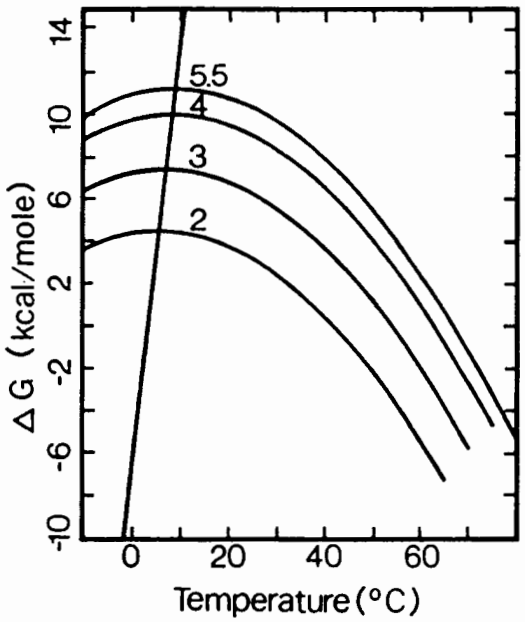

Fig. 6. Variation of stability curves with pH for a model protein with $\Delta\bar{H}(298)$ equal to 55.0 kcal/mole and a constant $\Delta\bar{C}_p$ of 2 kcal/deg mole. The model represents the denaturation of wild-type T4 lysozyme from pH 2 to pH 5.5. The solid straight line is the linear variation of the enthalpy with temperature.

THE LAW OF SMALL PERTURBATIONS ON PROTEIN STABILITY

We now inquire into the effect produced by a small chemical or physical change in the protein. Our original interest in this problem was derived from studies on the effect of mutations on protein stability,[29] but the result we present here may be applied to other perturbations on stability such as chemical modification, selective binding of ligand by the native or denatured state, changes in pH, or other solution conditions. The basic experiment is the comparison of two stabilities, as depicted in Fig. 7. The left part of the figure represents the unfolding of a protein in a reference state (some standard specification of pH, salt, and buffer composition). If the perturbation involves a reversible process, such as the uptake of ligands, changes in pressure, etc., then the upper and lower horizontal limbs of the figure also represent reversible processes to give a complete thermodynamic cycle. Such cycles

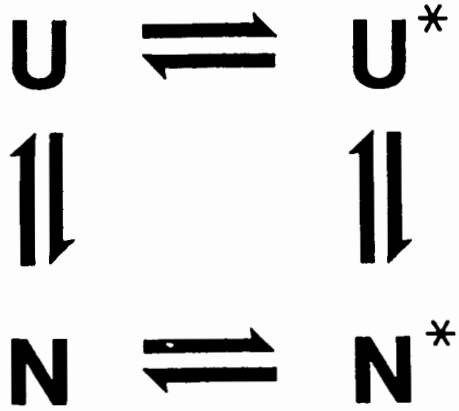

Fig. 7. The four thermodynamic states needed to compare a reference and perturbed protein. In this figure, N represents the folded state and U the unfolded state. The asterisks mark the states of the perturbed protein.

have been discussed theoretically[28] and have been observed experimentally.[23,24,30,31]

If, however, the perturbation is thé substitution of one amino acid for another, or an irreversible chemical modification, the two vertical limbs of the diagram are disconnected. There is no way of directly studying the thermodynamics of the mutational "reaction"

$$\text{Prot}(n = \text{Ala}) + \text{Gly} \rightarrow \text{Prot}(n = \text{Gly}) + \text{Ala}$$

$$\text{Prot}(n = \text{Ala}) + \text{X} - \text{Gly} - \text{Y} \rightarrow \text{Prot}(n = \text{Gly}) + \text{X} - \text{Ala} - \text{Y}$$

where the notation indicates an interchange between Ala in the n^{th} position of the protein with a free Gly in the first version, and a similar interchange with a Gly in the middle of a small peptide in the second. On the other hand, $\Delta\bar{G}$ and $\Delta\bar{G}^*$, the stabilities of the reference and perturbed protein, are well defined experimentally, and so is their difference, which we symbolize by g. In our notation

$$g = \Delta\bar{G}^* - \Delta\bar{G} \tag{17a}$$

$$h = \Delta\bar{H}^* - \Delta\bar{H} \tag{17b}$$

$$s = \Delta\bar{S}^* - \Delta\bar{S} \tag{17c}$$

g, as well as h and s, is a function of temperature and possibly the reference state as well as the perturbation.

It is possible to relate the change in melting temperature caused by a small perturbation to $g(T_g)$, the value of the perturbation free energy at the melting temperature of the protein in the reference state. This is demonstrated in Fig. 8. The slanting lines are the curves for $\Delta\bar{G}$ of the reference protein and $\Delta\bar{G}^*$ of the perturbed protein as they cross the abscissa at their melting temperatures

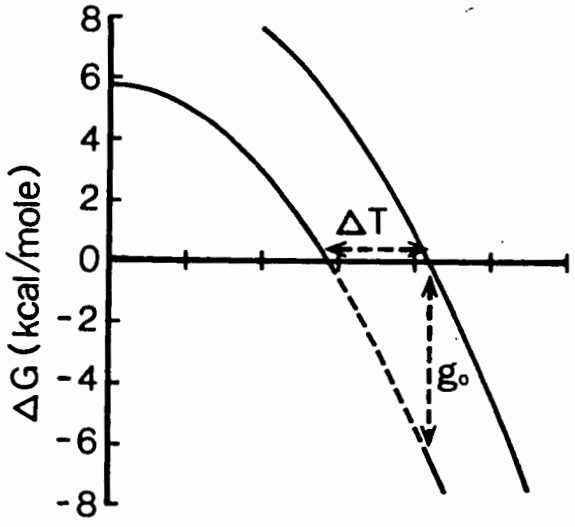

Fig. 8. The relationship between changes in melting temperature and changes in free energy. An explanation of the terms in this figure is given in the text.

T_g and T_g^*. As defined above, $g(T_g)$ is a vertical line connecting the two curves at T_g. Since $\partial\Delta\bar{G}/\partial T = -\Delta\bar{S}$, the slopes of the two curves at T_g and T_g^* are $-\Delta\bar{S}_g$ and $-\Delta\bar{S}_g^*$ respectively. If the stability curve for the perturbed protein can be approximated as a straight line, then from the figure,

$$\text{slope} = g(T_g)/(T_g^* - T_g) \quad \text{or} \quad \Delta T = g/\Delta\bar{S}_g^* \tag{18}$$

In the shorter notation on the right it is understood that g is evaluated at T_g and that ΔT is the difference between the melting temperatures of the mutant and reference proteins $(T_g^* - T_g)$. If we can further assume that the lines for $\Delta\bar{G}$ and $\Delta\bar{G}^*$ are parallel, then we can approximate $\Delta\bar{S}^*$ by $\Delta\bar{S}$ and write

$$\Delta T = g/\Delta\bar{S}_g = (gT_g)/\Delta\bar{H}_g \tag{19}$$

or in its most useful form

$$g = \Delta T(\Delta\bar{H}_g/T_g) \quad \text{or} \quad g = \Delta\bar{S}_g\Delta T \tag{20}$$

The utility of this formula is that if T_g and $\Delta\bar{H}_g$ are known for the protein in the reference state, then an estimate of g, the stabilization or destabilization of the perturbed state, can be obtained from a measurement of ΔT, which is the simplest measurement that one can make.

In our laboratory, data have been accumulated on a number of mutant proteins for which both ΔT and g are known (W. Becktel, W. Baase, and D. Muchmore, observations and papers in press). Table I compares the experimental values of g, which are of modest accuracy themselves, with the calculated values obtained with Eqs. (16) and (18). The agreement is good enough to justify the use of Eq. (16) for rough estimates of the effect of perturbations on stability. Equation (18) does not seem any more accurate and is not very useful. If one had enough information on the perturbed protein (mutant) to know $\Delta\bar{S}_g^*$, then g could be evaluated from $\Delta\bar{G}$ and $\Delta\bar{G}^*$

TABLE I
Comparison of Calculated and Measured Values of the Perturbation of Free Energies of Denaturation[a, b]

Protein	g_{exp}	g_{calc}	ΔT
Ile 3 → Cys ox	2.60	2.57	7.1
Ile 3 → Cys re	0.00	0.00	0.1
Met 6 → Ile	−1.38	−1.23	−3.4
Cys 54 → Thr	1.36	1.27	3.5
Arg 96 → His	−2.82	−2.83	−7.8
Ala 160 → Thr	−1.65	−1.99	−5.5
16H96[c]	−5.08	−4.69	−13.7

[a] The proteins were compared at pH 6. At this pH the melting point of the wild type is 65.5°C and its entropy of denaturation 362 e.u.

[b] g is in kcal/mole and ΔT in degrees.

[c] The notation 16H96 represents the double mutant of T4 lysozyme Met 6 → Ile/Arg 96 → His.

themselves. The main restriction on the formula is that g must be small and also that T_g must not be too close to T_s. When the latter is true, the maximum of the stability curve is close to T_g and the assumption of parallel straight lines crossing the abscissa becomes very poor. Since the stability curves are slightly curved downward in the region of interest, the line whose slope equals $g/\Delta T$ is in fact the chord of the actual $\Delta\bar{G}^*$ curve. Thus the most appropriate temperature at which to evaluate the entropy is between T_g and T_g^*. We have not attempted to use this refinement, although it may be justified with very accurate data.

The formula can also be derived analytically. The melting temperature of the perturbed protein is given by

$$T_g^* = \Delta\bar{H}_g^*/\Delta\bar{S}_g^* = \left[\Delta\bar{H}(T_g^*) + h\right]/\left[\Delta\bar{S}(T_g^*) + s\right] \tag{21}$$

If the numerator and denominator of this equation are expanded in powers of $(T_g^* - T_g)$, ignoring the small temperature dependence of h and s, then Eq. (16) is obtained as the lowest approximation and Eq. (17) as the next highest. Still higher terms are nonlinear. The only point of interest in the analytical derivation is that it shows that ΔT is intrinsically dependent on the free-energy factor g itself, and not directly on h or s. In Eq. (21), h and s are introduced as independent quantities, yet they come out automatically in the combination as $h - T_g s = g$.

We close this section with a discussion of a few particular forms of perturbation. This will relate our perturbation formula with several known results. We assume that the perturbation is dependent on a parameter α, which will be specified for each case. Since $\Delta\bar{S}_g$ and T_g are properties of the reference state and do not depend on the perturbation, we have, from Eq. (20),

$$\partial g/\partial\alpha = \Delta\bar{S}_g(\partial\Delta T/\partial\alpha) = \Delta\bar{S}(\partial T_g^*/\partial\alpha) \tag{22}$$

Case 1

The perturbation is the addition to the solution of a ligand that binds differentially to the folded and unfolded forms of the molecule. In this case Eq. (16) becomes

$$\partial\Delta T/\partial \ln L = -\left(RT_g/\Delta\bar{S}_g\right)\left(\partial \ln\left[\Sigma^u/\Sigma^f\right]/\partial \ln L\right) \tag{23}$$

and the parameter α is $\ln L$, where L is the activity of the ligand. We obtain the result

$$\frac{\partial\Delta T}{\partial \ln L} = -RT_g^2\Delta\nu/\Delta\bar{H}_g \tag{24}$$

where $\Delta\nu$ is the change in the number of ligands bound when the protein is unfolded. This formula is well established,[27,28] especially when pH is the variable, $\ln L = -2.3$ pH, and we obtain Eq. (15).

Case 2

The perturbation is the addition of a weakly bound solvent component and the effect is described as a virial expansion in the molality m_3, of the added component,[32] which might be a guanidinium salt or other perturbing reagent.[33] We are using the convention that component 1 is water, component 2 the protein, and component 3 is a small molecule such as a denaturing agent or salt. In this case

$$g = RT\Delta\beta_2 \tag{25}$$

where $\Delta\beta_2$ is the contribution of the interaction to $\Delta\bar{G}$ and $\alpha = m_3$. Thus we have

$$\left(\frac{\partial T_g^*}{\partial m_3}\right) = \frac{RT_g^2}{\Delta\bar{H}_g}\left(\frac{\partial\Delta\beta_2}{\partial m_3}\right) \equiv \frac{RT_g^2}{\Delta\bar{H}_g}\Delta\beta_{23} \tag{26}$$

If $\Delta\beta_{23}$ is proportional to m_3, $\Delta\beta_{23} = \Delta\beta^0 m_3$, which seems to be the case in a number of instances, then

$$\partial\Delta T/\partial m_3 = -RT^2\Delta\beta_{23}^0/\Delta\bar{H}_g \tag{27}$$

This supplies a way of evaluating the interaction parameter $\Delta\beta_{23}$ by studying changes in melting temperature. Since it can be shown that

$$\Gamma_{32} = \frac{-\Delta\beta_{23}m_3}{(1 + \beta_{33}m_3)} \tag{28}$$

where Γ_{32} is the thermodynamic binding parameter, $\beta_{33} = (\partial\mu_3/\partial m_3)_{\mu 2}$. This provides the connection with the stoichiometric binding formula, Eq. (24).

Case 3

The perturbation is an increase in pressure and α is the pressure itself. In this case

$$\int_{p_0}^{p}\Delta\bar{V}\,dp \tag{29}$$

and the formula becomes

$$\partial T_g^*/\partial P = \Delta V/\Delta\bar{S} \tag{30}$$

This is just the Clapyeron equation applied to a macromolecular transformation.

CONCLUSION

This paper has defined and discussed the stability curve of a protein or other molecule that undergoes a two-state transition, and has demonstrated its utility for the comparison of different proteins by evaluating the changes

induced in a protein by physical, chemical, or biological means. Apart from the section on the effect on varying $\Delta\bar{C}_p$, the results have been based on Privalov's model of constant $\Delta\bar{C}_p$. In the main the work has been concerned with the inferences that can be drawn by thermodynamic methods from the basic postulates of a two-state transition including the constancy of $\Delta\bar{C}_p$. Among these conclusions are the existence of low-temperature melting and of a temperature of maximum stability.

The three characteristic temperatures (T_g, T_s, and T_h) are of considerable heuristic value in comparing the properties of proteins with one another and in discussing changes in stability caused by changes in protein structure or environment. Simple and useful relations exist among these quantities and other thermodynamic quantities such as $\Delta\bar{C}_p$ and the maximum free energy.

Probably the most important new result for practical purposes is the formula relating ΔT, the change in melting temperature of a protein caused by a small alteration in protein structure or environment, to the change in stability relative to a reference protein at its melting temperature. The proportionality coefficient must be determined by a careful study of the transition of the reference protein to obtain its enthalpy of unfolding at T_g, but once this is done changes in stability may be estimated directly from ΔT. This is advantageous to laboratories that are not equipped to make careful studies of protein transitions. Much of the data that exists in the literature, especially for the effect of mutations, consists of simple melting-point determinations. Even without the determination of the proportionality constant, we have a rationale for ordering the stability of a series of mutants via their melting temperatures.

This research was supported by NSF grant PCM 8104339 and NIH grant GM 20195. The ideas presented here have been tested and improved by constant interaction with Bao-Lu Chen and F. Charlotte Schellman. We are especially indebted to Walter Baase, who forced us to think correctly about the problem of small perturbations, in order to provide a basis for the interpretation of his double-mutant work. We also thank Robert L. Baldwin for sharing his ideas and results on the hydrophobic effect with us prior to their publication.

References

1. Eisenberg, M. A. & Schwert, G. W. (1951) *J. Gen. Physiol.* **34**, 383–406.
2. Schellman, J. A. (1955) *Comp. Rend. Trav. Carlsberg Lab., Ser. Chim.* **15**, 230–259.
3. Brandts, J. F. (1969) in *Structure and Stability of Biological Macromolecules*, Timasheff, S. N. & Fasman, G.D., Eds., Marcel Dekker, NY, pp. 213–290.
4. Tanford, C. (1968) *Adv. Protein Chem.* **20**, 121–282.
5. Privalov, P. L. (1979) *Adv. Protein Chem.* **33**, 167–241.
6. Frank, H. S. & Evans, W. (1945) *J. Chem. Phys.* **13**, 507–532.
7. Kauzmann, W. (1959) *Adv. Protein Chem.* **14**, 1–64.
8. Brandts, J. F. (1964) *J. Am. Chem. Soc.* **86**, 4291–4301.
9. Shiao, D. F., Lumry, R. & Fahey, J. (1971) *J. Am. Chem. Soc.* **93**, 2024–2035.
10. Privalov, P. L. & Khechinashvili, N. N. (1974) *J. Mol. Biol.* **86**, 665–684.
11. Brandts, J. F. (1965) *J. Am. Chem. Soc.* **87**, 2759–2760.
12. Pace, C. N. & Tanford, C. (1968) *Biochemistry* **7**, 198–208.
13. Nojima, H., Ikai, A., Oshima, T. & Noda, H. (1977) *J. Mol. Biol.* **116**, 429–442.
14. Christensen, J. K. (1952) *Comp. Rend. Trav. Carlsberg Lab., Ser. Chim.* **28**, 37–169.
15. Brandts, J. F., Olivera, R. J. & Westort, C. (1970) *Biochemistry* **9**, 1038–1047.
16. Becktel, W. J., Baase, W. A., Wetzel, R. & Perry, J. J., (1986) *Biophys. J.* **49**, 109a.
17. Cho, K. C., Poon, H. J. & Choy, C. L. (1982) *Biochem. Biophys. Acta* **701**, 206–215.

18. Privalov, P. L., Griko, Yu. U. Venyaminov, S. & Kutyshenko, V. O. (1986) *J. Mol. Biol.* **190**, 487–498.
19. Baldwin, R. L. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 8069–8072.
20. Gill, S. J., Nichols, N. F. & Wadso, I. (1976) *J. Chem. Thermo.* **8** 445–452.
21. Dec, S. F. & Gill, S. J. (1984) *J. Sol. Chem.* **13**, 27–41.
22. Sturtevant, J. M. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 2236–2240.
23. Pfeil, W. & Privalov, P. L. (1976) *Biophys. Chem.* **4**, 23–32.
24. Pfeil, W. & Privalov, P. L. (1976) *Biophys. Chem.* **4**, 41–50.
25. Kirkwood, J. G. & Oppenheim, E. (1961) *Chemical Thermodynamics*, McGraw-Hill, NY.
26. Tanford, C. (1966) *Physical Chemistry of Biological Macromolecules*, John Wiley, NY.
27. Hermans, J. H., Jr. & Scheraga, H. A. (1961) *J. Am. Chem. Soc.* **83**, 3283–3292.
28. Schellman, J. A. (1975) *Biopolymers* **14**, 999–1018.
29. Schellman, J. A., Lindorfer, M., Hawkes, R. & Grutter, M. (1981) *Biopolymers* **20**, 1989–1999.
30. Pfeil, W. & Privalov, P. L. (1976) *Biophys. Chem.* **4**, 33–40.
31. Pace, C. N. & McGrath, R. (1980) *J. Biol. Chem.* **255**, 3862–3865.
32. Schellman, J. A. (1978) *Biopolymers* **17**, 1305–1322.
33. von Hippel, P. H. & Wong, K. Y. (1965) *J. Biol. Chem.* **240**, 3909–3929.

Received February 2, 1987
Accepted May 21, 1987