5. THE STABILITY OF FOLDED PROTEINS

We have now worked our way from the structure and chemistry of individual amino acids to the structures of folded proteins and their multimers. All along the way, the entropy of the peptide chain has fought against the adoption of a single folded conformation, but I've alluded to some of the compensating contributions of intermolecular forces that provide some enthalpic stabilization to a folded conformation and of course the hydrophobic effect, which uses the entropy of the solvent to drive the folding process. The question for this section is how much do (or could) each of these factors contribute to stability of soluble proteins? We'll also look at the special case of transmembrane proteins and explore how they are stabilized in a non-polar environment.

The Big Picture

C. Nick Pace at Texas A&M has done some great work on evaluating the various contributors to protein stability. Overall, for a 96-residue protein, ribonuclease Sa, he has estimated the free energy values contributing to folding in three rough categories.

- Loss of conformational entropy (+ 170 $^{kcal}/_{mol}$; this implies an average of 1.7 $^{kcal}/_{mol \bullet residue \bullet}$)
- The hydrophobic core (-100 $^{kcal}/_{mol}$)
- Hydrogen bonding (-80 $^{kcal}/_{mol}$)

Which means that, all told, the folded protein has a stability of about 10 $^{kcal}/_{mol}$ relative to the unfolded protein in solution.[1] Note that "hydrophobic core" here includes the enthalpic contribution of dispersion forces and the entropic contributions of the hydrophobic effect (solvent entropy).

The conformational cost of folding is, of course, the significant barrier to protein structure. I won't go into it any further here except to say that there are many conformational states available to each residue in the unfolded state that are no longer accessible in the folded state. That is what makes protein structure improbable. Our focus below is on what makes it possible. Before getting there, though, we need to cover some of the experimental methods that are used to evaluate stability.

Experimental Determination of Protein Stability

The goal of this lecture is to describe methods for obtaining structure-stability relationships in protein structure. In previous lectures we've discussed the size and shape side of things, but have not addressed in any quantitative way the interplay of enthalpy and entropy in shaping proteins. Studies directed at such information typically require systems that operate in a two-state equilibrium. In general, the two states will be an unfolded (U) state in equilibrium with a native (N), folded state:

$$N \leftrightarrows U; \qquad K_{unfolding} = [U]/[N] \qquad \text{(Eq. T.1)}$$

1 Pace, C Nick; Grimsley, Gerald R; and Scholtz, J Martin (December 2009) Protein Stability. In: Encyclopedia of Life Sciences (ELS). John Wiley & Sons, Ltd: Chichester. DOI: 10.1002/9780470015902.a0003002.pub2

As will be discussed later, it is not always easy to set up experiments in which the folded an unfolded states are in a two-state equilibrium, but for our initial purposes, we will discuss two techniques that provide thermodynamic data derived from such equilibrium studies.

van't Hoff Analysis

Equilibrium constants are temperature dependent. At a simple level (naively, and incorrectly, assuming that ΔH and ΔS are constant within the temperature range of interest), the relationship can be captured in the van't Hoff equation (Eq. T.3), which is readily derived from the relationship between equilibrium constants and free energy (Eq. T.2):

$$\ln(K_{eq}) = -\Delta G^{\circ}/RT \qquad \text{(Eq. T.2)}$$

$$\ln(K_{eq}) = -\frac{\Delta H^{\circ}}{R}\left(\frac{1}{T}\right) + \frac{\Delta S^{\circ}}{R} \qquad \text{(Eq. T.3)}$$

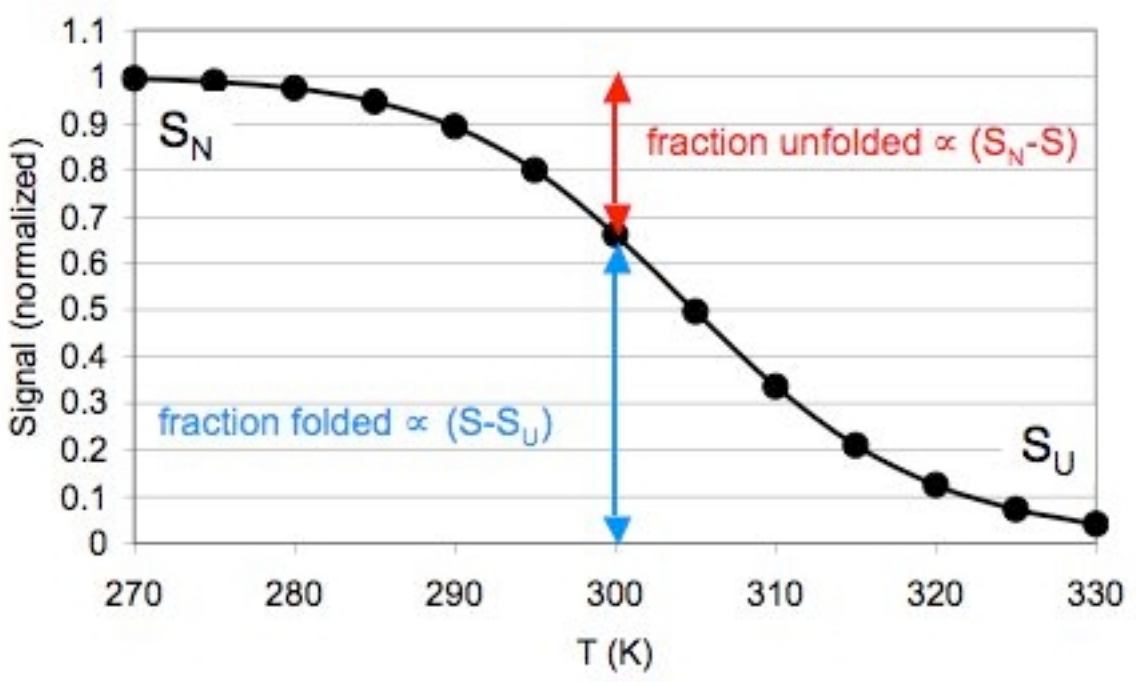


Figure T.1. Temperature dependent equilibrium study. Note that the maximum signal is obtained for the folded (native) form and the unfolded form gives the minimal signal in this instance. Experimental data is never this clean, and often baselines must be obtained to cope with temperature dependence of the signal that is independent of the chemical system in question.

By obtaining equilibrium data for a **two-state** process at multiple temperatures, one may plot the natural logarithm of the equilibrium constants vs. inverse temperature (measured in Kelvins) to obtain ΔH° and ΔS° from the slope and intercept, respectively. It is common in structural equilibrium studies to focus on a spectroscopic signal (such as circular dichroism or fluorescence

emission intensity) that correlates to a structural change. Such a signal (S) will have two extreme values, S_U and S_N, for the unfolded and native forms. In the course of measuring the structural equilibrium at multiple temperatures, the signal will evolve as the stability of one state gradually degrades and the other begins to predominate (Figure T.1) The equilibrium constant for folding (for example) may then be expressed as follows (Eq. T.4):

$$K_{unfold} = \frac{[U]}{[N]} = \frac{\text{fraction unfolded}}{\text{fraction folded}} = \frac{S_N - S}{S - S_U} \qquad \text{(Eq. T.4)}$$

Working with the data in Figure T.1, one can obtain reasonable measurements of the equilibrium constant between 290 K and 320 K, where $0.1 \leq K_{unfold} \leq 10$. Those boundaries are chosen (somewhat arbitrarily) because there are measurable amounts of both N and U present at each temperature. Outside those ranges, measurement becomes more difficult. Table T.1 contains retabulated data for K_{unfold} that may be used to prepare a van't Hoff plot.

Table T.1. The equilibrium constant for unfolding of a protein between 280-320 K, using the plot in Figure T.1.

Temp (K)	**S**	**K $(S_N-S)/(S-S_U)$**	**1/T (K^{-1})**	**ln(K)**
290	0.89	0.12	0.00345	-2.12
295	0.80	0.25	0.00339	-1.38
300	0.66	0.51	0.00333	-0.67
305	0.50	1.0	0.00328	0.017
310	0.34	2.0	0.00323	0.68
315	0.21	3.8	0.00317	1.33
320	0.12	7.0	0.00313	1.95

The plot of ln(K) against 1/T then will give a slope of $-\Delta H^\circ/R$ and a intercept of $\Delta S^\circ/R$ (Figure T.2).

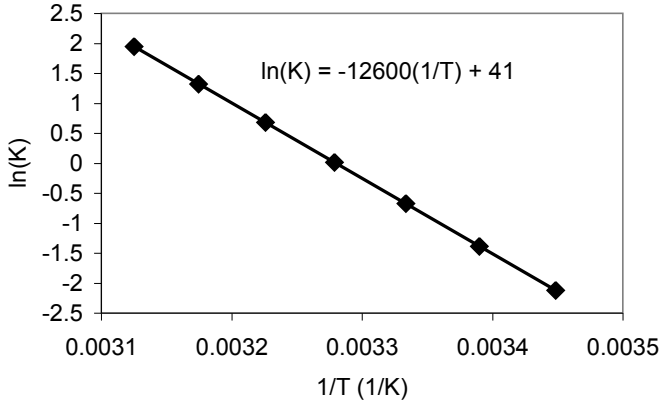


Figure T.2. van't Hoff plot arising from data shown in Figure T.1. The slope of -12600 K indicates a ΔH° of unfolding of +25 kcal/mol and a ΔS° of unfolding of +82 cal/mol•K.

Working with denaturants

Sometimes, however, a two-state equilibrium cannot be obtained under conditions that might be deemed physiological. This is particularly a problem for protein unfolding studies. Often, unfolded proteins spontaneously and irreversibly aggregate (think of an egg white hitting a frying pan). Thus, it is sometimes necessary to study protein unfolding under **denaturing** conditions, where the unfolded protein is both stable and soluble.

Figure T.3. Structures of two denaturing agents (chaotropes), urea on the left and guanidinium chloride on the right.

The two principle denaturants used in these studies are urea and guanidinium chloride (Figure T.3). These **denaturing** species are **chaotropes**, meaning that they diminish the hydrophobic effect and enhance the solubility of non-polar molecules in aqueous solution. For example, benzyl alcohol is 2.5 times more soluble in 8 M urea than water. The chemical basis for chaotropic behavior is that these species disrupt hydrogen-bonding networks in water absent any non-polar solute. Thus, upon addition of a non-polar species, there is not enthalpic drive for water to form clathrates in order to maintain H-bonding networks around the solute (since it didn't have any to begin with), and there is not entropic cost to the water molecules. Likewise, chaotropic agents interact favorably with non-polar surface area. The vdW interactions between urea and non-polar hydrocarbons are stronger than between the hydrocarbon moieties themselvs, providing an enthalpic benefit for the dissolution of nonpolar solutes in concentrated solutions of urea and guanidinium. Proteins **denature** under such circumstances and adopt an unfolded conformation.

A common method of employing chaotropes in unfolding studies is to measure the equilibrium constant for unfolding as a function of [denaturant] instead of temperature. At each concentration, and equilibrium constant is found, and used to determine a free energy of unfolding at that denaturant concentration, which may be plotted vs. denaturant concentration to exploit the following relationship (Eq. T.5, assuming urea):[2]

$$\Delta G_{unfold}^{[urea]} = \Delta G_{unfold}^{H2O} - m[urea] \qquad \text{Eq. T.5}$$

As urea concentration increases, the free energy of unfolding becomes more spontaneous, and thus more negative. The value of the slope, -**m,** is generally a good indication of the amount on non-polar surface area exposed upon unfolding. The greater the dependence of unfolding on urea concentration, the more positive **m** value and the more non-polar surface area becoming exposed. The truly useful value to arise from this type of study is the free energy of unfolding in water, which is the y-intercept obtained by plotting equation T.5 (Figure T.4). To use this data to obtain $\Delta H°$ and $\Delta S°$ for unfolding, simply perform the study at multiple temperatures.

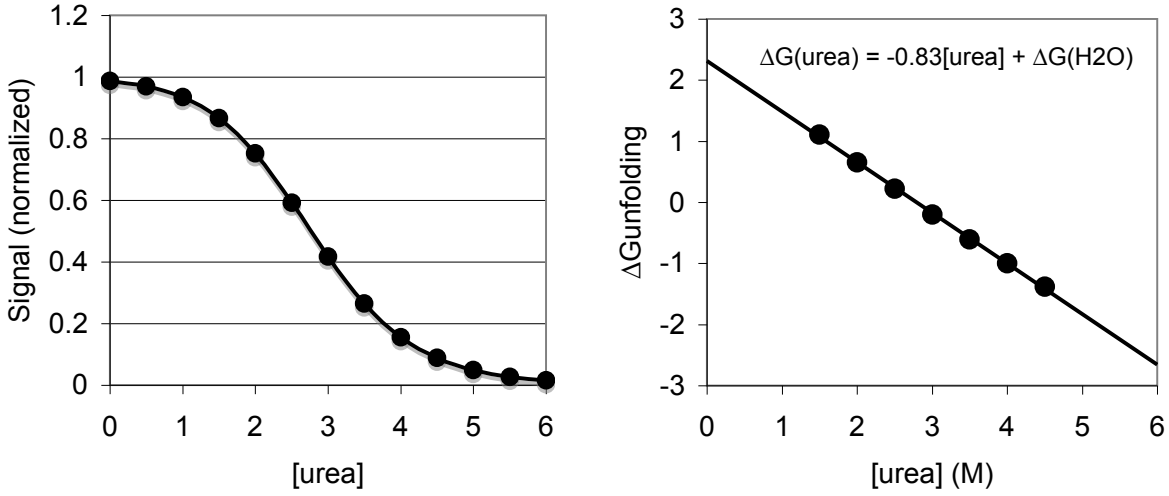


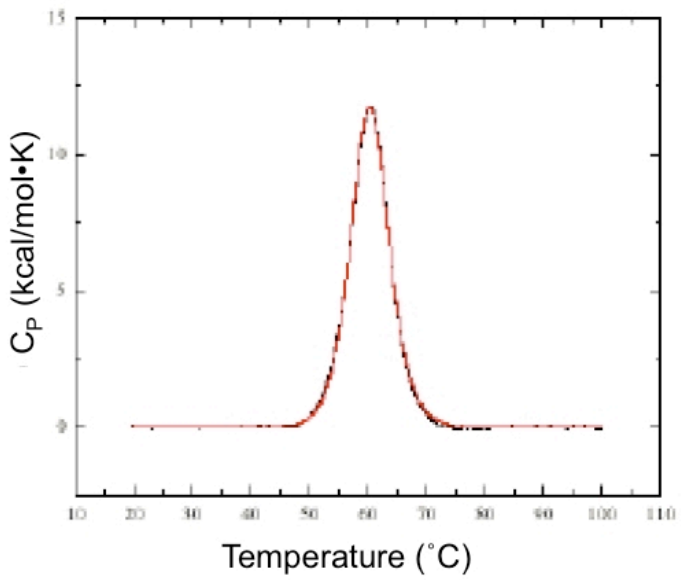
Figure T.4. Urea denaturation study. At left, a plot of some spectroscopic signal vs. [urea]. The analysis is performed as with Figure T.1, but the K_{unfold} values obtained at each concentration of urea are used to calculate ΔG_{unfold} for that [urea]. That is replotted at right to obtain a linear relationship between free energy of unfolding vs. [urea]. In this instance, the ΔG_{unfold} in water is +2.5 kcal/mol and m is 0.83.

Differential Scanning Calorimetry

Differential scanning calorimetry (DFC) is an alternate technique for obtaining thermodynamic data involves measuring the amount of heat that must be applied to a sample to raise the temperature 1°C. This is the heat capacity of the sample. Most transitions from the folded to unfolded state are endothermic. As one raises the temperature from a low starting point, one approaches a temperature at which the equilibrium constant between the folded and denatured state is roughly unity (~40°C in Figure T.5). This is the so-called **melting temperature**, $\mathbf{T_m}$**.** The heat capacity of a protein solution becomes larger during the melting transition because most of the heat being applied to the solution is going into the unfolding of the protein, rather than into raising the temperature of the solution.[2]

In a simple DFC experiment, a solution of a folded protein sample is gradually warmed, while monitoring the heat input to the system. As the protein unfolds, the endotherm is registered by the increased flow of heat, leading to a peak in the plot of heat capacity (C_P) vs. temperature. To obtain ΔH_{unfold}, one integrates the area under the curve, relative to the baseline, and the excess heat absorbed is the enthalpy of unfolding at the melting temperature (Eq. T.6). Similarly, one can obtain the ΔS_{unfold} by integrating the excess heat applied, divided by temperature (Eq. T.7). The melting temperature, in which ΔG_{unfold} is zero, is the ratio of ΔH to ΔS (Eq. T.8).

[2] This is analogous to melting ice. One may apply a great deal of heat to a mixture of ice and water, but it will remain at 0°C until all the ice has melted. The applied heat simply goes into converting ice to water.



$$\Delta H^\circ = \int_{T_1}^{T_2} \Delta C_P dT \qquad \text{(Eq. 5.6)}$$

$$\Delta S^\circ = \int_{T_1}^{T_2} \frac{\Delta C_P}{T} dT \qquad \text{(Eq. 5.7)}$$

$$T_m = \frac{\Delta H^\circ}{\Delta S^\circ} \qquad \text{(Eq. 5.8)}$$

Figure T.5. Idealized DSC scan of a protein unfolding event. The sharp increase in C_P (the heat capacity – or the amount of energy needed to raise the sample temperature by a given amount) is due to absorption of heat by the unfolding process. In reality, changes in heat capacity before and after unfolding and buffer effects would generally cause a more complicated plot.

Comparison of van't Hoff and Calorimetric Results

Often one will see ΔH_{vH} and ΔH_{cal} explicitly distinguished from one another. This is because the methods have the potential to yield differing results based on the system being studied. Where one has a simple one-to-one, two state equilibrium (Eq. T.1), the two values will be identical. But often, there are additional interactions in one of the two states that complicate matters. For example, if the folded protein exists as a dimer, the two-state equilibrium will be:

$$N_2 \Leftrightarrow 2U \qquad K_{unfold} = \frac{[U]^2}{[N_2]} \qquad \text{(Eq. T.9)}$$

However, one might not necessarily know that the folded state is dimeric and assume that the ΔH_{vH} obtained reflects unfolding of a single chain. Calorimetric measurements allow direct determination of the enthalpy absorbed on a per chain basis, since one has a direct measurement of the heat absorbed and one knows the total number of chains. If ΔH_{cal} is smaller than ΔH_{vH}, that is evidence that some additional interactions are taking place in the folded state that give a larger enthalpy of unfolding than is detected calorimetrically.

Sources of Stability

As noted earlier, protein stability is derived chiefly through formation of a hydrophobic core and secondarily through polar interactions such as hydrogen-bonding. These stabilizing features compensate the folded polypeptide chain for its loss of conformation entropy, and then provide just a little extra stabilization to achieve stability.

The Hydrophobic Core

Pace[3] estimates that each methylene (CH_2) group buried in the hydrophobic core contributes about 1.2 kcal/mol in free energy to the stability of a protein. That number is obviously an average, and individual measurements vary considerably. The problem is that most experiments approach the issue by removing methyl(ene) groups from the core which has two effects: (1) a loss of vdW interactions and hydrophobic effect associated with the missing group and (2) a deformation of the protein's conformation as it attempts to compensate for that loss. Item "1" is what one hopes to study. Item "2" is a complication that clouds the measurements.

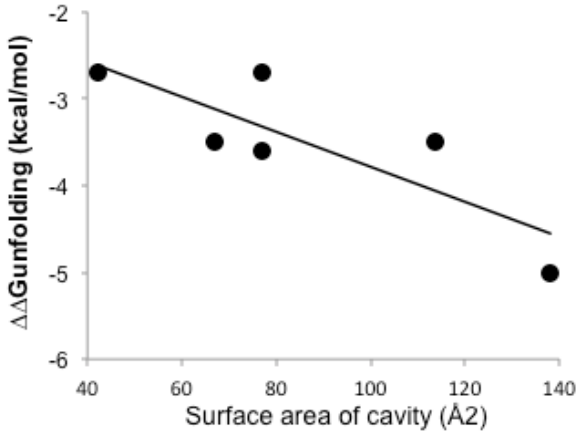


Figure T.6. Effect of cavity creation on the stability of T4 lysozyme. Leucine and phenylalanine were mutated to alanine in order create internal cavities. Cavity size varies due to rearrangement of side chains. A plot of $\Delta\Delta G_{unfolding}$ vs. the cavity surface area shows that each $Å^2$ of cavity surface leads to about a 20 cal/mol destabilization of the protein.

[3] Pace et al. (2004) "Protein structure, stability and solubility in water and other solvents" *Phil. Trans. R. Soc. London* **359**, 1225–1235

A thorough study of substitutions to the core was performed by Brian Matthews' lab at the University of Oregon.[4] Working with the protein T4 lysozyme, the group introduced a variety of cavities into the hydrophobic core by mutating leucine and phenylalanine residues to alanine. Rather than trust no structural changes would accompany these deletions, x-ray crystal structures were performed, with cavities varying in size from 50 to 150 $Å^3$ with a Leu→Ala mutation. Thermal stability was tested for each mutant by van't Hoff analysis. Enthalpic decrease in stability ranged from roughly 30 to 45 kcal/mol with these mutations, but comparisons are complicated because the values are measured at different melting temperatures. A sounder comparison could be made by investigating ΔG at a specific temperature (51.8 °C here), leading to the plot in figure T.6. The roughly linear relationship of stability to cavity surface area (area that has lost vdW interactions in the core, and has not benefitted from the hydrophobic effect in folding) shows that each 1 $Å^2$ of cavity surface area leads to 20 cal/mol destabilization of the protein. If a methylene group is assumed to have a surface area of about 40 $Å^2$, then each contributes about 0.8 kcal/mol to the stability of a folded protein.

As an aside to this study, the Leu99 to Ala mutation opened a 150 $Å^3$ hole in the hydrophobic core and lowered the melting temperature to 36°C from the native 52°C. In one of those experiments that begs the question "why not", they exposed the Leu99Ala mutant to benzene, which fits nicely into the cavity left behind. The melting temperature increased by 6°C and the stability of the protein recovered by 2 kcal/mol. Nature abhors a cavity.

From here I want to cover some of Pace's stuff. In 2004 Royal Soc paper he provides a breakdown of enthalpy/entropy to hphobic effect, claiming about 1 kcal/mol is entropic while 0.2 is enthalpic.

Enthalpic contributions

Hydrogen Bonding

Numerous hydrogen bonds form within folded proteins as a result of secondary structure elements, but there are also a smaller number of H-bonds that form between side chains that contribute to the stabilization of tertiary structure. These are less common but can play a role.

Fill-in here: Fersht and barnase

Then: Kelly and PIN

Salt Bridges

Electrostatic interactions can be a contributor to protein stability, especially in the context of a hydrogen bond. Ion-ion interactions between acidic (anionic) and basic (cationic) residues provide

[4] Eriksson et al. (1991) "Response of a Protein Structure to Cavity-Creating Mutations and Its Relation to the Hydrophobic Effect" *Science* **255**, 178-183.

some enthalpic stability to a protein and have been shown to stabilize α-helices when the residues are positioned one turn apart on the same face of a helix.[5]

However, the contribution such "salt bridges" make to protein stability should not be over-emphasized. In a fairly surprising result, the Sauer lab showed that a network of H-bonded residues (Arg – Glu – Arg) buried in the Arc repressor protein could successfully be replaced with a number of roughly isosteric nonpolar residues, with *no* loss of stability. Basically, adding residues to the hydrophobic core is as valuable to the protein as adding charge-charge interactions.

[5] Marqusee and Baldwin (1987) Helix stabilization by Glu-Lys+ salt bridges in short peptides of de novo design. *PNAS* **84**, 8898.