

Equilibrium constants and free energies in unfolding of proteins in urea solutions

(multiple binding/virtual equilibrium constants/denaturation/cooperativity)

IRVING M. KLOTZ*

Department of Chemistry and Department of Biochemistry, Molecular Biology and Cell Biology, Northwestern University, Evanston, IL 60208-3113

Contributed by Irving M. Klotz, October 8, 1996

ABSTRACT A novel thermodynamic approach to the reversible unfolding of proteins in aqueous urea solutions has been developed based on the premise that urea ligands are bound cooperatively to the macromolecule. When successive stoichiometric binding constants have values larger than expected from statistical effects, an equation for moles of bound urea can be derived that contains imaginary terms. For a very steep unfolding curve, one can then show that the fraction of protein unfolded, \bar{B} , depends on the square of the urea concentration, U , and is given by

$$\bar{B} = \frac{A_1^2 e^{\lambda n \bar{B}} U^2}{1 + A_1^2 e^{\lambda \bar{B}} U^2}.$$

A_1^2 is the binding constant as $\bar{B} \rightarrow 0$, and λ is a parameter that reflects the augmentation in affinities of protein for urea as the moles bound increases to the saturation number, n . This equation provides an analytic expression that reproduces the unfolding curve with good precision, suggests a simple linear graphical procedure for evaluating A_1^2 and λ , and leads to the appropriate standard free energy changes. The calculated ΔG° values reflect the coupling of urea binding with unfolding of the protein. Some possible implications of this analysis to protein folding *in vivo* are described.

The effects of high concentrations of urea in aqueous solution on the properties of proteins have intrigued investigators since early in this century (1–3). In the same era, the reversibility of denaturation by various reagents, including urea, was discovered (4, 5). Some decades later when spectroscopic probes responsive to the conformational states of proteins had been developed, the quantitative dependence of denaturation on urea concentration could be ascertained. Such data provided a foundation for a thermodynamic approach to the unfolding process.

A representative unfolding curve for a protein [bacterial histidine-containing protein, HPr (6)] in aqueous solutions of increasing urea concentration is illustrated in Fig. 1. Since the transition is reversible, it has become customary to view each point as an equilibrium position of the interconvertible native, N, and denatured forms, D, respectively, and to define an equilibrium constant $K_{N-D} = (D)/(N)$. This constant can then be used to compute a standard free energy of unfolding, ΔG_{D-N}° (7–12).

As is apparent from Fig. 1, the K_{N-D} is extremely sensitive to the concentration of urea in the aqueous solvent. An increase in urea concentration of less than 2-fold (from 3.5 to 6 M) changes K_{N-D} by more than a 100-fold. This is surprising and perplexing, for urea is a very innocuous solute, very compatible with water. At the gross or macroscopic level, aqueous urea solutions show no exceptional properties. True, the dielectric increment of urea, +2.7 (13), implies that in a 3–6 M solution it raises the dielectric constant of the water from near 80 to the

range of 88–96. This contrasts with the effects of most other common organic solutes in water, which reduce the dielectric constant slightly. On the other hand, even 1 M glycine in water raises the dielectric constant to 100, and 5 M β -alanine in water increases the dielectric constant to 250 (13). Yet, protein–ligand binding studies (14) show that the latter two solutes, in contrast to urea, stabilize the native conformation of the protein rather than facilitate denaturation.

The very steep rise in Fig. 1 is reminiscent of ligand–receptor behavior when there is very strong cooperativity between successively bound molecules. The strong cooperativity in the urea/HPr system is particularly evident if one compares the experimentally observed curve with that for an “ideal” transformation in which successive steps of ligand binding are energetically intrinsically identical and which manifests a halfway point at the same concentration of urea as does HPr unfolding. Below 2 M urea, the free energy of unfolding of the HPr is so unfavorable that the transformation is essentially undetectable, even though the ideal system attains a conversion near 30%. Above 2 M urea, a steep and increasing upward rise appears so that the conversion for unfolding quickly overtakes and surpasses that for the ideal system. The observed unfolding of the protein approaches completion at a urea concentration (≈ 6.5 M) that achieves only 60% conversion for the ideal transformation.

Solutions of 4–8 M urea contain 25–50%, by weight of solute; the molar concentration of urea reaches 10^6 -fold that of the protein. It behooves us, therefore, to examine the behavior of the protein if urea molecules are bound with high cooperativity. This can be approached from a purely phenomenological, classical, thermodynamic perspective, not dependent on a two-state conformational model for the protein.

Binding Equations for Strongly Coupled Successive Steps

The increasing effectiveness of urea (U) in the unfolding of proteins (P) can be formulated quantitatively in classical thermodynamic terms with a stepwise, stoichiometric representation of ligand binding to the protein:

$$\begin{aligned} P + U &= PU; & K_1 &= \frac{(PU)}{(P)(U)} \\ PU + U &= PU_2; & K_2 &= \frac{(PU_2)}{(PU)(U)} \\ &\dots & & \\ &\dots & & \\ &\dots & & \\ &\dots & & \\ PU_{n-1} + U &= PU_n; & K_n &= \frac{(PU_n)}{(PU_{n-1})U}. \quad [1] \end{aligned}$$

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: HPr, histidine-containing protein.
*e-mail: chemdept@chem.nwu.edu.

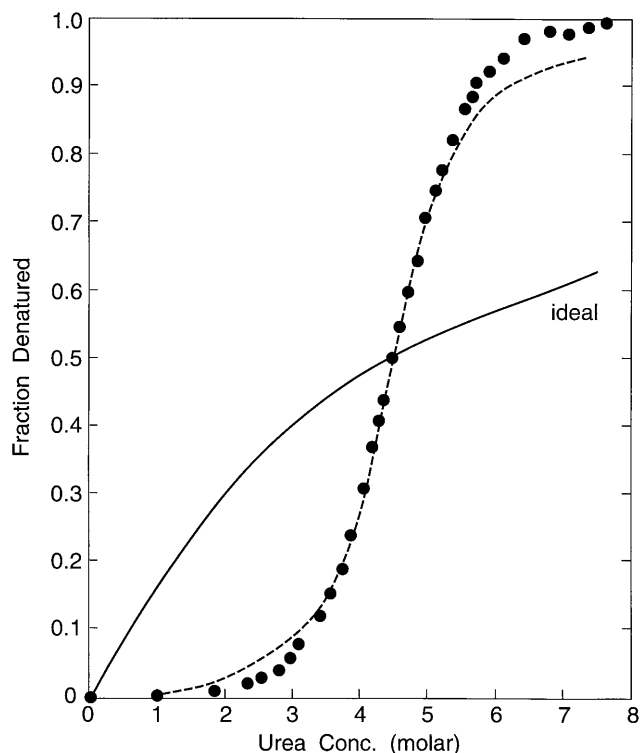


FIG. 1. Unfolding curve for HPr, bacterial histidine-containing protein, in aqueous urea solutions. Experimental data from ref. 6. The broken line through the points presents the curve calculated from Eq. 27. The "ideal" curve is that for a transformation in which successive steps of ligand binding are energetically intrinsically identical.

The subscript n denotes the moles of bound urea at saturation of the receptor. The moles B of ligand, urea, bound per mole of protein can be related to the concentration of free ligand, (U), by the equation (15, 16):

$$B = \frac{K_1(U) + 2K_1K_2(U)^2 + \cdots + j(K_1K_2 \cdots K_j)(U)^j + \cdots}{1 + K_1(U) + K_1K_2(U)^2 + \cdots + (K_1K_2 \cdots K_j)(U)^j + \cdots} \quad [2]$$

where K_j (which is always real and positive) represents the stoichiometric equilibrium constant for the successive stoichiometric step, j ,

$$PU_{j-1} + U = PU_j; \quad K_j = \frac{PU_j}{(PU_{j-1})(U)} \quad [3]$$

An alternative algebraic expression for B , following from the fundamental theorem of algebra (16) applied to Eq. 2, is

$$B = \frac{K_\alpha(U)}{1 + K_\alpha(U)} + \frac{K_\beta(U)}{1 + K_\beta(U)} + \cdots + \frac{K_\omega(U)}{1 + K_\omega(U)} + \cdots \quad [4]$$

The coefficients $K_\alpha \cdots K_\omega$ have the dimensions of equilibrium constants and can be combined to yield functions of the stoichiometric equilibrium constants $K_1, K_2 \cdots K_j$. However, they do not correspond to any particular step in the successive equilibria of Eq. 1 nor to the binding of ligand by any specific site on the receptor, so they may be viewed as *virtual* equilibrium constants.

When the uptake of one mole of ligand enhances the affinity of the receptor for the succeeding mole—i.e., when the ratio of successive stoichiometric binding constants is greater than statistical (in which case the binding is frequently called "cooperative"), the constants $K_\alpha, K_\beta \cdots K_\omega$ are complex numbers appearing in conjugate pairs. For a bivalent receptor, for example, which becomes saturated when two moles of ligand are bound,

$$K_\alpha = a + bi; \quad K_\beta = a - bi, \quad [5]$$

where $i = \sqrt{-1}$ and the coefficients a and b are real numbers. Alternatively, in exponential notation,

$$K_\alpha = Ae^{i\theta}; \quad K_\beta = Ae^{-i\theta}, \quad [6]$$

where $A = \sqrt{a^2 + b^2}$ and $\theta = \arctan(b/a)$. Eq. 4 for such a bivalent system becomes

$$B = \frac{Ae^{i\theta}(U)}{1 + Ae^{i\theta}(U)} + \frac{Ae^{-i\theta}(U)}{1 + Ae^{-i\theta}(U)}. \quad [7]$$

Recalling the relation

$$e^{ix} + e^{-ix} = 2 \cos x \quad [8]$$

we can convert Eq. 7 into

$$B = \frac{A(2 \cos \theta)(U) + 2A^2(U)^2}{1 + A(2 \cos \theta)(U) + A^2(U)^2}. \quad [9]$$

It should be noted in passing that one can show readily (15) that

$$A^2 = K_1K_2. \quad [10]$$

$$2A \cos \theta = K_1. \quad [11]$$

$$\frac{A}{2 \cos \theta} = K_2. \quad [12]$$

In the general case, a multivalent system where saturation of receptor is reached when n moles of ligand are bound, and the accentuation of successive stoichiometric equilibrium constants continuously exceeds statistical values, then (15, 16),

$$B = \sum_{\ell=1}^{n/2} \frac{A_\ell(\cos \theta_\ell)(U) + 2A_\ell^2(U)^2}{1 + A_\ell(2 \cos \theta_\ell)(U) + A_\ell^2(U)^2}. \quad [13]$$

Originally there were n terms in this series, but they were condensed to $n/2$ pairs by steps corresponding to that used in the transformation of Eq. 7 to Eq. 9.

Application to Unfolding by Urea

As Fig. 1 illustrates, curves for protein unfolding in urea solutions are remarkably steep. The accentuations in successive stoichiometric equilibrium constants must be very large. For a bivalent system, for example, K_2 must be much larger than K_1 . From Eqs. 11 and 12 we see that in the limit K_2/K_1 approaches infinity for $\theta = \pi/2$. Under these circumstances, Eq. 9 becomes

$$B = \frac{2A^2(U)^2}{1 + A^2(U)^2}. \quad [14]$$

This corresponds to uptake of ligand in coupled pairs.

For a multivalent system, the corresponding binding equation is

$$B = 2 \sum_{\ell=1}^{n/2} \frac{A_1^2 U^2}{1 + A_1^2 U^2}. \quad [15]$$

This corresponds to progressive uptakes of coupled pairs of ligand molecules.

If one has a cluster of binding sites that in the absence of ligand are essentially identical, one can partition the free energy of binding into three terms (17):

$$\Delta G^\circ = \Delta G_{\text{intrinsic}} + \Delta G_{\text{statistical}} + \Delta G_{\text{interaction}}. \quad [16]$$

The first term on the right-hand side expresses the affinity for a ligand at any site isolated from any influence of other sites on the receptor. The second term takes account of the relative number of open and of occupied sites, which can affect the extent of binding of ligand even when there is no interaction between sites. The third term specifies the impact of interactions between occupied sites on the resultant affinity for ligand. A more explicit expression is

$$\Delta G_j^\circ = -RT \ln K_j = -RT \ln K_{\text{intrinsic}} - RT \ln(\text{statistical factor}) + \Delta G_{\text{interaction}}, \quad [17]$$

which leads to

$$K_j = K_{\text{intrinsic}} \left(\frac{n-j+1}{j} \right) e^{\frac{-\Delta G_{\text{interaction}}}{RT}}. \quad [18]$$

For a facilitative interaction, which leads to increases in affinity constants K_j , the free energy of interaction must be negative, so the exponent in Eq. 18 becomes a positive number. If the facilitative interaction continues to add to the affinity progressively with increasing moles bound ligand, B , then the exponent will increase with B .

It has been shown (18, 19) that when n , the number of sites on the receptor is large, equations for moles bound, B , can be reduced to simpler algebraic form. For the binding of pairs of urea ligands at the $n/2$ receptor sites of a protein, one can replace Eq. 15 by the expressions

$$B = 2 \left[\frac{n}{2} \frac{A_1^2 e^{\lambda B} U^2}{1 + A_1^2 e^{\lambda B} U^2} \right] \quad [19]$$

$$\bar{B} = \frac{B}{n} = \frac{A_1^2 e^{n\lambda \bar{B}} U^2}{1 + A_1^2 e^{n\lambda \bar{B}} U^2}, \quad [20]$$

where A_1^2 is the binding constant for the uptake of the first pair of urea ligands and λ is a factor that reflects the enhancement in affinities as B (or \bar{B}) increases.

For convenience in evaluating the parameters A_1^2 and λ for a specific unfolding curve, one can make the following algebraic rearrangements.

$$\bar{B}(1 + A_1^2 e^{n\lambda \bar{B}} U^2) = A_1^2 e^{n\lambda \bar{B}} U^2. \quad [21]$$

$$\frac{\bar{B}}{U^2} = A_1^2 e^{n\lambda \bar{B}} [1 - \bar{B}]. \quad [22]$$

$$\ln \left(\frac{\bar{B}}{1 - \bar{B}} \frac{1}{U^2} \right) = \ln A_1^2 + \lambda n \bar{B}. \quad [23]$$

The ratio $[\bar{B}/(1 - \bar{B}) \times (1/U^2)]$ and may be viewed as the operative equilibrium constant, K_{oper} , during unfolding of protein in the presence of urea. Eq. 23 suggests that $\ln K_{\text{oper}}$ be plotted as a function of fractional conversion \bar{B} . In the limit of decreasing values of \bar{B} ,

$$\lim_{\bar{B} \rightarrow 0} (\ln K_{\text{oper}}) = \ln A_1^2. \quad [24]$$

If the interaction parameter, λ , remains constant with increase in \bar{B} , then $n\lambda$ will be fixed and the data will follow a straight line, the slope of which is $n\lambda$. Also, as \bar{B} approaches 1,

$$\lim_{\bar{B} \rightarrow 1} (\ln K_{\text{oper}}) = \lim_{\bar{B} \rightarrow 0} (\ln K_{\text{oper}}) + \lambda n. \quad [25]$$

Free energy changes corresponding to K_{oper} can be computed from the equation

$$-\Delta G^\circ = RT \ln A_1^2 + RT n \lambda \bar{B}. \quad [26]$$

Comparison with Experiment

Fig. 2 presents a transform of the unfolding curve of Fig. 1 into the format specified by Eq. 23. The experimental points shown, encompassed within the range $0.15 < \bar{B} < 0.85$, fall along a straight line, in conformance with Eq. 23 when λ is a constant.

Experimental points for \bar{B} below 0.15 or above 0.85 are not displayed in Fig. 2. Since the value of \bar{B} depends on a difference between two experimental readings of a physical quantity (circular dichroic θ , fluorescence intensity, or absorbance), the error therein is larger as \bar{B} approaches the asymptotes where \bar{B} or $1 - \bar{B}$ tends to zero. If the points for $\bar{B} < 0.15$ are added to Fig. 2, they fall progressively below the line as $\bar{B} \rightarrow 0$. Such a trend suggests a systematic error in the choice of an appropriate value for the physical quantity assigned to the fully folded protein, which is used in every computation of \bar{B} . Similarly it has been found that points for $B > 0.85$ trace out a curve rising progressively above the straight line shown. Such a trend is likely due to a systematic error, in this region, in the choice of the value for the physical quantity assigned to the fully unfolded protein.

Using the data for the interval $0.15 < \bar{B} < 0.85$, the range of highest reliability, one can extrapolate the line observed to $\bar{B} = 0$ and thereby evaluate A_1^2 . For bacterial HPr, this mean value of the operative binding constant is 0.83×10^{-2} . Using Eq. 25, one finds λn is 3.5.

Thus one can now return to the explicit expression for \bar{B} as a function of U^2 , Eq. 20:

$$\bar{B} = \frac{0.83 \times 10^{-2} e^{3.5 \bar{B}} U^2}{1 + 0.83 \times 10^{-2} e^{3.5 \bar{B}} U^2}. \quad [27]$$

This provides an analytic representation of the graphical information. The curve corresponding to Eq. 27 is displayed in Fig. 1 as a broken line (- - -). It is apparent that in the range $0.15 < \bar{B} < 0.85$, the agreement with the published experimental observations (6) is excellent, as would be expected from the close fit in Fig. 2. At the peripheries, $\bar{B} < 0.15$ and $\bar{B} > 0.85$, there are systematic deviations. As mentioned earlier, these may be manifestations of defects in choices of limiting values in the asymptotic approaches to fully folded and completely unfolded protein. Alternatively, the theoretical analysis may need further refinement.

Corresponding to the operative equilibrium constant,

$$K_{\text{oper}} = A_1^2 e^{\lambda n \bar{B}}, \quad [28]$$

a standard free energy change ΔG° has been defined as:

$$-\Delta G^\circ = RT \ln A_1^2 + RT n \lambda \bar{B}. \quad [29]$$

It is apparent that ΔG° varies with the extent of unfolding. In the limit of zero urea concentration, where $\bar{B} = 0$, the limiting value of ΔG° is $2840 \text{ cal} \cdot \text{mol}^{-1}$. This corresponds to $-RT \ln A_1^2$, the standard free change associated with the uptake of the first

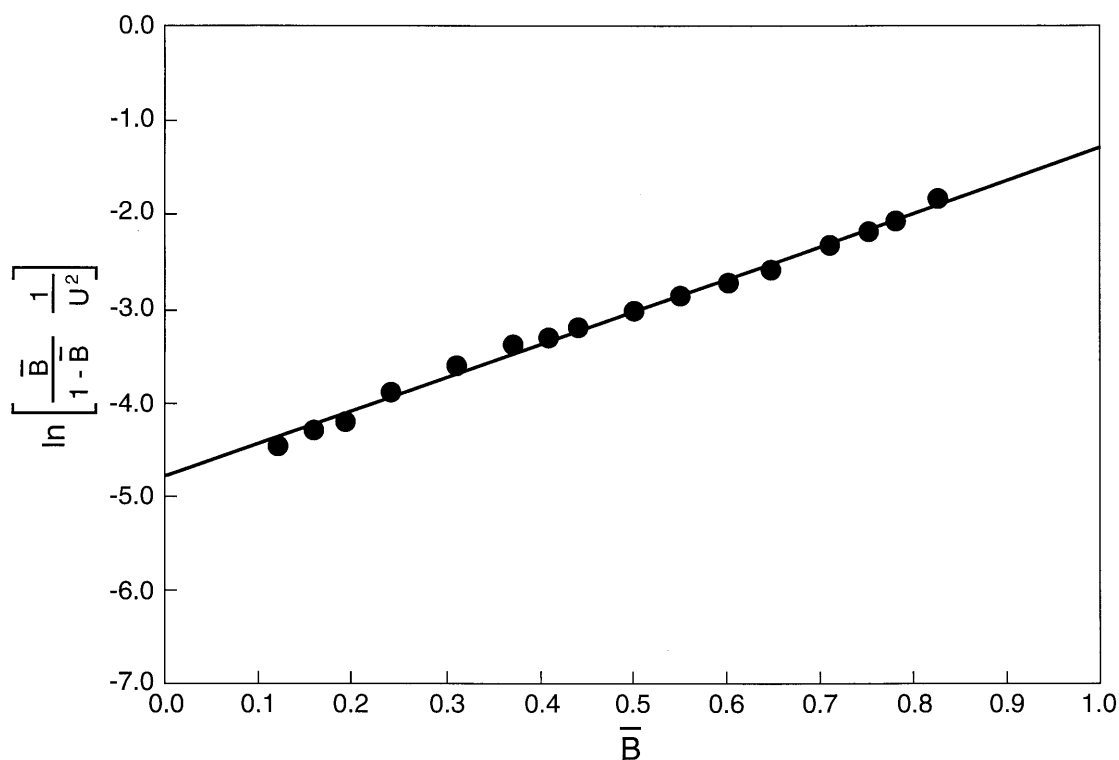


FIG. 2. Unfolding data for HPr protein in urea solutions presented in terms of the logarithm of the operative binding constant, K_{oper} , as a function of the fraction unfolded, \bar{B} .

pair of urea ligands by the empty sites of the protein macromolecule in solution in pure water.

Implications

In the usual treatment of unfolding data in terms of a single $N = D$ equilibrium, the ΔG_{N-D}° obtained (6, 12) for bacterial HPr by extrapolation to zero urea concentration is 4210 cal·mol⁻¹, substantially different from the 2840 cal·mol⁻¹ computed in this paper. The discrepancy is only an apparent one, for one must recognize that the two ΔG° values refer to different processes. That for the $N = D$ equilibrium is the free energy change for the conversion of native protein in its standard state ("hypothetical" 1 molar) to fully unfolded protein in its standard state ("hypothetical" 1 molar):

$$N \left(\begin{array}{c} \text{std} \\ \text{state} \end{array} \right) = D \left(\begin{array}{c} \text{std} \\ \text{state} \end{array} \right) \quad \Delta G^\circ = 4210 \text{ cal}\cdot\text{mol}^{-1}. \quad [30]$$

On the other hand, that calculated by the procedure in this paper is for the transformation

$$N \left(\begin{array}{c} \text{std} \\ \text{state} \end{array} \right) + 2U \left(\begin{array}{c} \text{std} \\ \text{state} \end{array} \right) = DU_2 \left(\begin{array}{c} \text{std} \\ \text{state} \end{array} \right) \\ \Delta G^\circ = 2840 \text{ cal}\cdot\text{mol}^{-1}. \quad [31]$$

These are different chemical reactions so it is not surprising that they are associated with different values of ΔG° .

Nevertheless, it is still appropriate to compare ΔG° values for wild-type and mutant proteins.

For example, for the HPr protein, a mutant, S31A, in which the Ser-31 of the wild type was replaced by an Ala, shows an unfolding curve similar to that illustrated in Fig. 1 but displaced slightly toward lower urea concentrations (6). At zero urea concentration, the ΔG° value for S31A is 2640 cal·mol⁻¹. Thus $\Delta(\Delta G^\circ)$ is -200 cal·mol⁻¹ (2640 - 2840). The $\Delta(\Delta G^\circ)$ evaluated from the midpoint of unfolding curves (11, 12) is

essentially the same, for the interaction parameter, the slope of the line in Fig. 2, is the same for S31A as for HPr. If the $\Delta(\Delta G^\circ)$ is computed on the basis of the two-state assumption, Eq. 30, a value of -460 cal·mol⁻¹ is obtained (6). From either result one can see that the mutant protein is slightly more disposed toward unfolding.

The change in ΔG° , 200 cal·mol⁻¹, cannot be ascribed to a weakening of certain bonds in the protein (20), for it may be instead a reflection of increased affinity of the receptor for urea molecules. Binding of solute molecules and opening of bonds in the protein are coupled during unfolding by urea, and from the experiments at hand it is not possible to deconvolute the net result into the independent contributions of different forms of interactions. To achieve that insight one needs electromagnetic sensors in addition.

The analysis in this paper has focused on the *unfolding* of the protein. Since the curve in Fig. 1 refers to a reversible conformational change, the *folding* process must also be a cooperative one, along the reverse pathway. Viewed from that perspective, it is apparent that the folding increases from 10% to 90% for a drop in concentration of urea from about 5.5 to 3.5 molar. This is a very sharp transition for a relatively small fractional drop in ligand concentration, to about 0.65 the starting value. If the binding of ligand were ideal, its concentration would have to change from about 40 molar to 0.4 molar; that is, the drop would have to be to about 0.01 of the starting value, to change the folding of the protein from 10% to 90%. Thus, with cooperative ligand binding the conformation of the macromolecule can be manipulated over a wide range by a small perturbation in ligand concentration.

As has been known for decades, solutes other than urea can drive protein unfolding, most of them at far lower concentrations. Thus it has been shown that reversible denaturation of hemoglobin by salicylate (5) has a midpoint at 0.3 M concentration, about 1/15 that of the usual midpoint in a urea-driven unfolding curve. At, and below, 0.3 M in aqueous solution, salicylate produces no unusual effects on the gross, macro-

scopic character of the solvent. It seems likely that it effects unfolding by being bound to the protein.

This perspective suggests some heuristic speculations about what may control protein folding *in vivo*. If an appropriate ligand in the cytoplasm were bound to the polypeptide chain as it emerged from the ribosome, the chain could be maintained in an unfolded conformation. When the binding is cooperative, then a relatively small drop in ligand concentration could swing the protein conformation to a folded conformation. Such a change in ligand concentration could be effected by auxiliary factors (such as chaperones) capable of binding ligand competitively or of releasing ligand by forming an alternative complex with the protein. These coupled interactions could serve to regulate conformational changes in the macromolecule and the transition from unfolded to native form *in vivo*.

1. Ramsden, W. & Chavasse, N. G. (1913) *Trans. Faraday Soc.* **9**, 93–96.
2. Anson, M. L. & Mirsky, A. E. (1929) *J. Gen. Physiol* **13**, 121–132, 133–143.
3. Hopkins, F. G. (1930) *Nature (London)* **126**, 328–330, 383–384.
4. Northrop, J. H. (1932) *J. Gen. Physiol* **16**, 323–337.
5. Anson, M. L. & Mirsky, A. E. (1934) *J. Gen. Physiol.* **17**, 393–398, 399–408.
6. Hammen, P. K., Scholtz, J. M., Anderson, J. W., Waygood, E. B. & Klevit, R. E. (1995) *Protein Sci.* **4**, 936–944.
7. Schellman, J. A. (1955) *C. R. Trav. Lab. Carlsberg Ser. Chim.* **29**, 223–229.
8. Schellman, J. A. (1987) *Biopolymers* **26**, 549–559.
9. Aune, K. & Tanford, C. (1969) *Biochemistry* **8**, 4586–4590.
10. Pace, C. N. (1986) *Methods Enzymol.* **131**, 266–280.
11. Matouschek, A., Mathews, J. M., Johnson, C. M. & Fersht, A. R. (1994) *Protein Eng.* **7**, 1089–1095.
12. Scholtz, J. M. (1995) *Protein Sci.* **4**, 35–43.
13. Cohn, E. J. & Edsall, J. T. (1943) *Proteins, Amino Acids and Peptides* (Reinhold, New York), pp. 144–146.
14. Klotz, I. M. & Luborsky, S. W. (1959) *J. Am. Chem. Soc.* **81**, 5119–5124.
15. Klotz, I. M. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 7191–7194.
16. Klotz, I. M. (1997) *Ligand-Receptor Energetics: A Guide for the Perplexed* (Wiley, New York).
17. Klotz, I. M., Walker, F. M. & Pivan, R. B. (1946) *J. Am. Chem. Soc.* **68**, 1486–1490.
18. Scatchard, G. (1949) *Ann. N.Y. Acad. Sci.* **51**, 660–672.
19. Thompson, C. J. & Klotz, I. M. (1971) *Arch. Biochem. Biophys.* **147**, 178–185.
20. Mirsky, A. E. & Pauling, L. (1936) *Proc. Natl. Acad. Sci. USA* **22**, 439–447.