the
abdus salam
international centre for theoretical physics

SMR: 1098/12

WORKSHOP ON THE STRUCTURE OF BIOLOGICAL MACROMOLECULES

(16 - 27 March 1998)

"Interactions that Stabilize Protein Folding"

presented by:

Robert L. BALDWIN
Stanford University
School of Medicine - Beckman Center
Stanford, CA 94305-5307
U.S.A.

Lecture 1
R. L. Baldwin
March 23, 1998

Interactions that Stabilize Protein Folding

A. Overview of Factors Involved

1. 2-state folding, small proteins:
   - ΔG(unf) measurable from K ($U \rightleftharpoons N$)
   - $\Delta\Delta G$(unf) measurable for mutants
   - ΔH(unf) measurable by calorimetry
   - hydrogen exchange measures stability of individual peptide NH protons.
2. ΔG(unf) is a balance between a few major factors
   - (1) favorable:
     - the hydrophobic interaction
     - specific side chain interactions
   - (2) unfavorable:
     - conformational entropy
     - burial of polar groups
   - (3) question mark:
     - peptide H-bonds?
3. Specific side chain interactions include:
   - H-bonds
   - salt bridges (or ion pairs)
   - charge-helix dipole interactions
   - aromatic ring-charged histidine
4. Disulfide bonds stabilize folding:
   - stability depends on the effective concentration of the cysteine residue
5. Striking aspects of protein stability:
   - cold denaturation
   - heat-stable proteins from extreme thermophiles
   - marginal stability of isolated α-helices and reverse turns
   - inferences about stabilizing interactions from frequencies in protein structures

B. Measurement of the thermodynamics of unfolding

1. The 2-state model ($N \rightleftharpoons U$)
   - a. Validity of the model depends on the absence of intermediates.
   - b. Standard tests for the absence of intermediates are:
     - (i) use two different probes and test if they give the same transition curve;

(ii) using calorimentry (see below),
compare the measured ΔH (ΔH_{cal}) with the van't Hoff ΔH (ΔH_{vH}) calculated from the temperature dependence of the equilibrium constant.

2. Unfolding transition curves

a. To determine the transition curve, let the measured property (e.g. fluoresence or CD) be y and calculate the fraction of native protein (f_N) by

$$f_N = (y-y_U)/(y_N-y_U) \quad (1)$$

b. Then the equilibrium constant for folding (K=[N]/[U]) is given by

$$K = f_N/(1-f_N) \quad (2)$$

c. The free energy of folding is

$$\Delta G = RT \ln K \quad (3)$$

d. When a denaturant (urea or GdmCl (guanidinium chloride) is used to cause unfolding, ΔG in most cases varies linearly with C

$$\Delta G = \Delta G(0) + mC \quad (4)$$

where C is the molar concentration of the denaturant and m (or the "m-value") is slope of the plot.

e. The property y, which is used to measure the unfolding transition curve, is assumed to vary linearly with C (the denaturant concentration). The "baseline" dependences of y_N and y_U on C are determined by combining eqns (1-4) into one equation, and then fitting y versus C by least squares: see M.M.Santoro & D.W. Bolen (1988) Biochemistry 27, 8063-8068.

f. A good "how to do it" chapter is given by C.N.Pace, B.A. Shirley and J.A.Thomson in "Protein Structure: A Practical Approach", ed. by T.E.Creighton, IRL Press, 1989.

3. Theory of the linear extrapolation method

3. a. In the linear extrapolation method (LEM), ΔG is plotted against the molar concentration C of the denaturant and extrapolated linearly back to C=0 to find ΔG in water, in the absence of denaturant. How denaturants interact with proteins (and

why they interact more strongly with unfolded proteins than with native proteins) has been investigated in different ways, for example by measuring the heat (ΔH) of interaction of the protein and denaturant: see G.I.Makhatadze and P.L.Privalov (1992) J. Mol. Biol. 226, 491-505.

3. b. There are some subtle aspects of the interaction of denaturants with proteins, especially in the case of GdmCl; the problem has been analyzed in detail by J.A.Schellman (1990) Biophy. Chem. 37, 121-140.

3. c. A key point is that the interaction in most cases follows the weak interaction model, not the site binding model. (An example of an interaction that follows the site binding model is the well-known combination of protons with conjugate bases to form acids.) In the weak interaction model, the interacting species k simply perturbs the chemical potential of the protein (species i), rather than combining chemically with it. Thus, if the chemical potential of the protein (i) is written

$$\mu_i = \mu_i^\circ + RT \ln y_i C_i + RT \beta_i \qquad (5)$$

where the chemical potential μ_i is another way of writing the partial molar free energy Gi and where $RT \beta_i$ is the excess free energy resulting from the interaction of the protein (i) with the denaturant (k). The term β_i can be expressed by the first term of a series expansion.

$$\beta_i = \beta_{ik} C_k + \ldots \qquad (6)$$

With $\beta_i = \partial\beta_{ik}/\partial C_k$. At low protein concentrations the protein activity coefficient can be set equal to 1. At equilibrium between U and N, $\mu_N = \mu_U$ and this relation, together with eqns (5) and (6), gives

$$-RT \ln K = \Delta G(0) + RT\, C_k(\beta_{Nk} - \beta_{Uk}) \qquad (7)$$

This is the same as eqn (4) above, which provides the basis for the linear extrapolation method, with

$$m = RT\,(\beta_{Nk} - \beta_{Uk}) \qquad (7a)$$

$$\Delta G(0) = \mu^\circ_N - \mu^\circ_U \qquad (7b)$$

C. The Hydrophobic Interaction

1. a. A hydrophobic molecule is insoluble in water but soluble in nonpolar solvents. Thus, mercury, which is insoluble in both solvent systems, is not classified as hydrophobic.

1. b. The study of interfacial tension reveals that hydrophobic molecules are in fact attracted to water (a drop of hexane spreads over a water surface), but water is more strongly attracted to itself (C.Tanford, PNAS 76, 4175; 1979). The reason is that the H-bonds of water provide about 3-4 times more strength of attraction between surfaces than the van der Waals (or dispersion force) interactions. The latter are about equally strong, per unit area, between water and hydrocarbon and hydrocarbon with itself.

2. a. The nature of the H-bonded structure of water remains hotly controversial today, despite the large number of studies that have been made. Models fall roughly into two classes; (1) those in which water contains a mixture of normal H-bonds and broken H-bonds, and the fraction broken increases with temperature; (2) those in which all possible H-bonds are made but many are distorted, and the degree of distortion increases with temperature. The best-known model of the latter class is the "random network" model of S.A.Rice and M.G.Sceats, J. Phys. Chem. 85, 1108; 1981. It is possible to fit all the thermodynamic properties of water, including the third derivatives of the basic thermodynamic variables, to this model without using arbitrary adjustable parameters: A.R.Henn & W.Kauzmann, J. Phys. Chem. 93, 3770; 1989.

2. b. In his discussion of the structure of water, C.Tanford ("The Hydrophobic Effect", John Wiley & Sons; 1986; Chapter 5) makes these two points.

   (1) The radial distribution function of liquid water, which gives the distribution of distances between molecules and the occupancy at each spacing, shows that two of three spacings between next-nearest neighbors is consistent with the icelike tetrahedral model but one spacing is not.

   (2) Comparison between the heat of sublimation of ice and its heat of melting indicates that 15% of the H-bonds in ice are broken in water at 0°, when these numbers are interpreted by the first model above.

3. a. Kauzmann's original proposal for quantitating the hydrophobic interaction by studies of transfer energies (W.Kauzmann, Adv. Protein. Chem. 14, 1; 1959) is still used today.

3. b. The transfer energies are used to define the proportionality factor between solvent-

accessible nonpolar surface area and Gibbs energy of transfer, as proposed by C.Chothia (Nature 248, 338; 1974) R.B.Herman (J. Phys. Chem. 71, 2754; 1972) and C.Tanford and coworkers (PNAS 72, 2925; 1974). The coefficient given by Chothia is about 25 cal mol^{-1} $A^{\circ -2}$.

4. a. The surface area of a molecule depends on how it is defined. The standard definition used today is that given by B.-K.Lee and F.M.Richards, JMB 55, 379; 1971. The area is obtained by rolling a water molecule over the test substance and taking its surface area as the locus of points available to this probe.

4. b. Use of the Lee & Richards algorithm is straightforward when the structure of the molecule is well-defined. To obtain the contribution of the hydrophobic interaction to the ΔG for protein folding, it is necessary to define the structure of the unfolded protein. The extended chain is often used but this practice has been criticized and other estimates of the actual solvent-exposed surface of an unfolded protein have been suggested: see G.D.Rose and coworkers, Biochemistry 34, 16245; 1995.

5. a. A basis for the proportionality between the transfer energy and the surface area is given by the cavity model, which is an accurate relation for a macroscopic cavity.

$$\Delta G = A\gamma \qquad (1)$$

Here ΔG is the work of making a cavity in the solvent, A is the surface area of the cavity and γ is the surface tension of the liquid, when the cavity is empty. When this model is used as a rough guide to the transfer energy for a hydrocarbon equilibrated between water and a nonpolar solvent, the transfer energy is $NA(\gamma_1-\gamma_2)$ where N is Avogadro's number, A is the accessible surface of the hydrocarbon solute, γ_1 is the surface tension of water (72 erg cm^{-2} at 20°) and γ_2 that of the nonpolar solvent (about 25 erg cm^{-2}). This equals 67 cal mol^{-1} $A^{\circ -2}$.

5. b. B.Honig and coworkers (Science 252, 106; 1991) suggest that in considering microscopic cavities, the curvature of the cavity should be included in computing the transfer energy because the accessibility to other water molecules of a water molecule probe at the surface increases as the cavity becomes smaller. Taking account of this effect leads them to change the proportionality factor between transfer energy and surface area from 25 to 45 cal mol^{-1} $A^{\circ -2}$.

6. The amount of nonpolar surface area buried upon folding of single-domain proteins is proportional to the molecular weight. R.S.Spolar et al. (Biochemistry 31, 3947; 1992) give the proportionality coefficient as 0.64: see equation 19.

6. a. Thus, if the average molecular weight per residue is 110, 70 $A^{\circ -2}$ of nonpolar surface is buried per residue during folding. This corresponds to 1.8 kcal/res of free energy gained per residue, using Chothia's (1975) value of 25 cal/$A^{\circ 2}$. For a 100-residue protein, burial of nonpolar surface provides 180 kcal/mol of free energy driving folding.This is by far the largest single factor favoring folding.

6. b. The largest factor opposing folding is the conformational entropy change, which has been estimated by E.Freire and coworkers as 6 e.u., or 1.8 kcal/residue at 298K (Proteins 25, 143; 1996).

6. c. Thus, these two main factors nearly cancel if these estimates are correct, and the observed free energy of folding (typically -5 to -10 kcal/mol) represents other factors such as sidechain interactions and peptide hydrogen bonds.

Figures

1. From C. N. Pace (1990) Tibtech 8, 93
   Upper: urea unfolding curve for RNase A measured by fluorescence
   Lower: thermal unfolding curves for RNase T1, with ΔG plotted against temperature
2. Upper: from P. L. Privalov Transition temperature (T_m) plotted again pH for metmyoglobin (circles, with arms), RNase A (triangles), cytochrome c (circles), α-chymotrypsin (diamonds), and hen lysozyme (squares),
   Lower: from Santoro & Bolen (1988)
3. Upper: dependence of the enthalpy of unfolding of T4 lysozyme on temperature; from R. Hawkes et al. (1984) J. Mol. Biol. 175, 195.
   Lower: calorimetric measurements of the enthalpy of unfolding of hen lysozyme at various pH values (Privalov and coworkers).
4. Upper: clathrate water structure surrounding n - $C_4H_9S^+F^-$ (from R. E. Dickerson and I. Geis, "The Structure and Action of Proteins", 1969).
   Lower: transfer free energy versus accessible surface area of amino acid side chains and some hydrocarbons: from C. Chothia (1975). Nature 254, 304.

Fig. 1

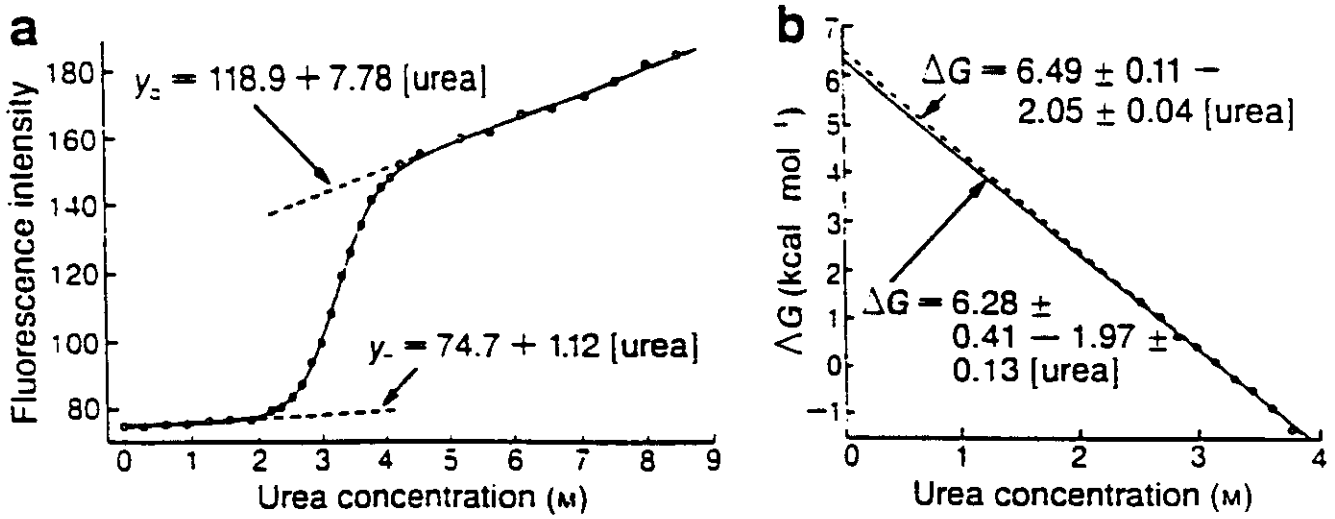
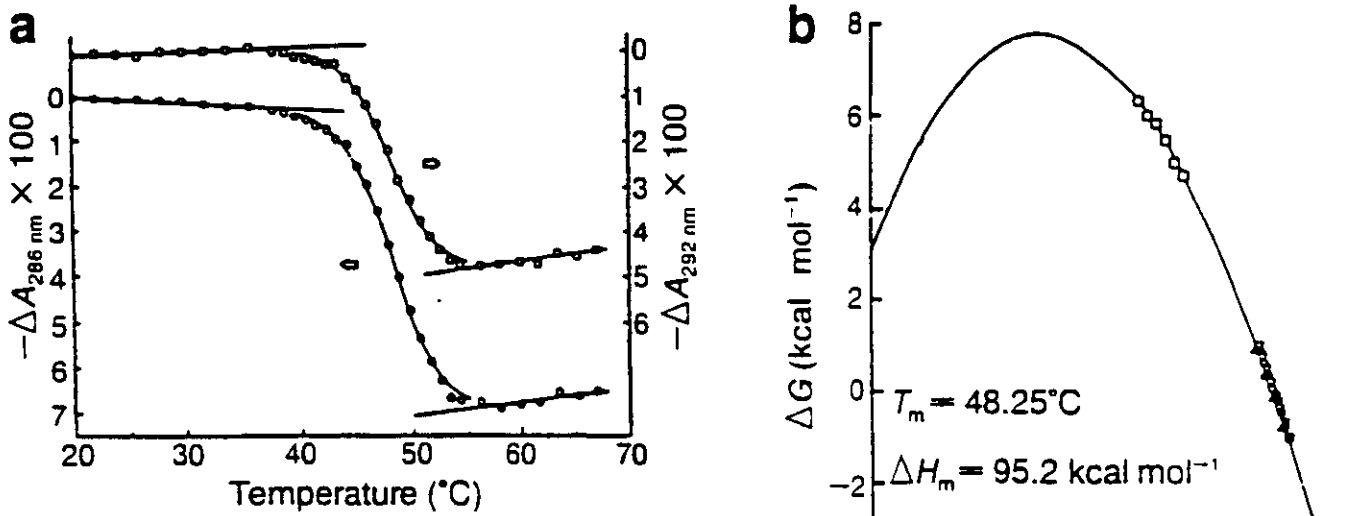
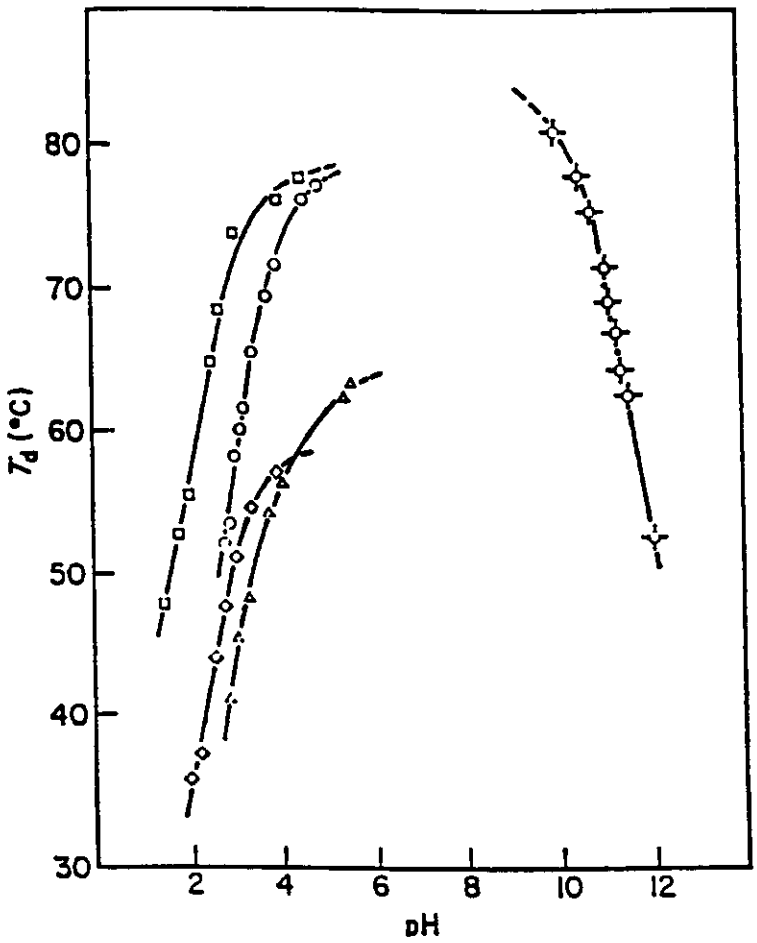
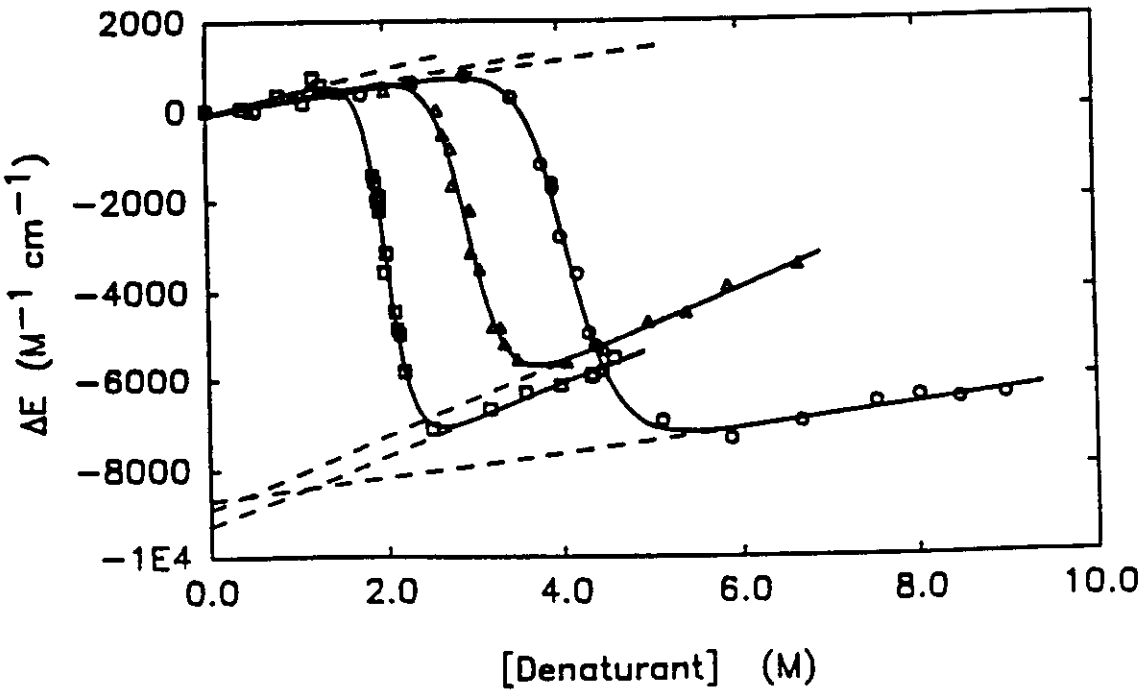


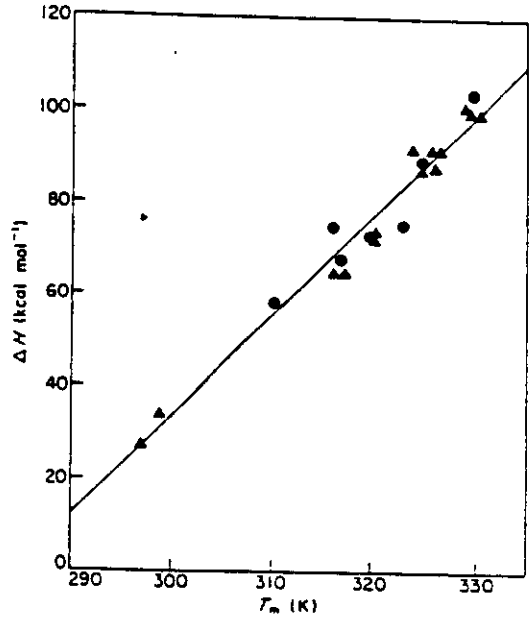
Fig. 2



(a) Thermal unfolding curves for RNase T1 in 30 mM MOPS buffer, pH 7.0. The difference in absorbance between RNase T1 at 25°C and at the temperature


80
70
60
50
40
30
T_d (°C)
2
4
6
8
10
12
pH


2000
0
−2000
−4000
−6000
−8000
−1E4
ΔE (M^{-1} cm^{-1})
0.0
2.0
4.0
6.0
8.0
10.0
[Denaturant] (M)



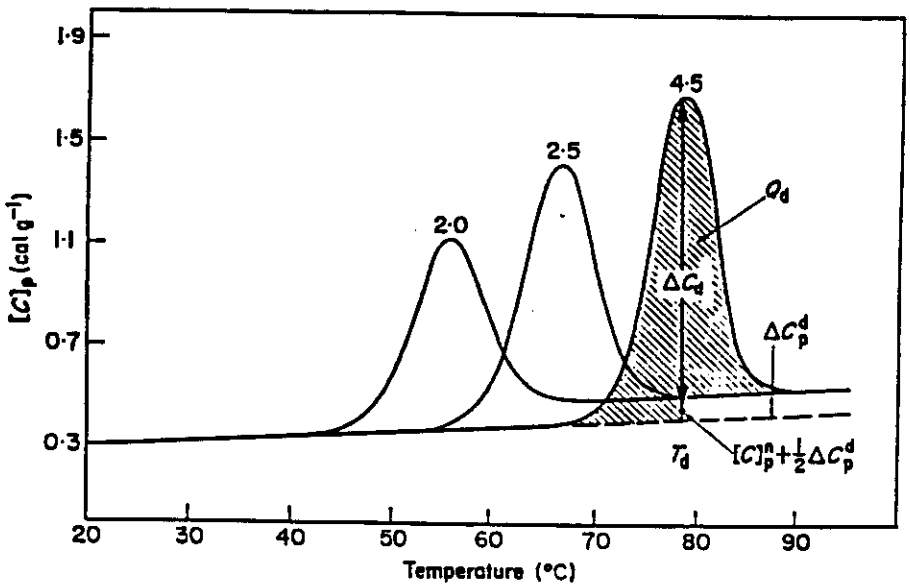
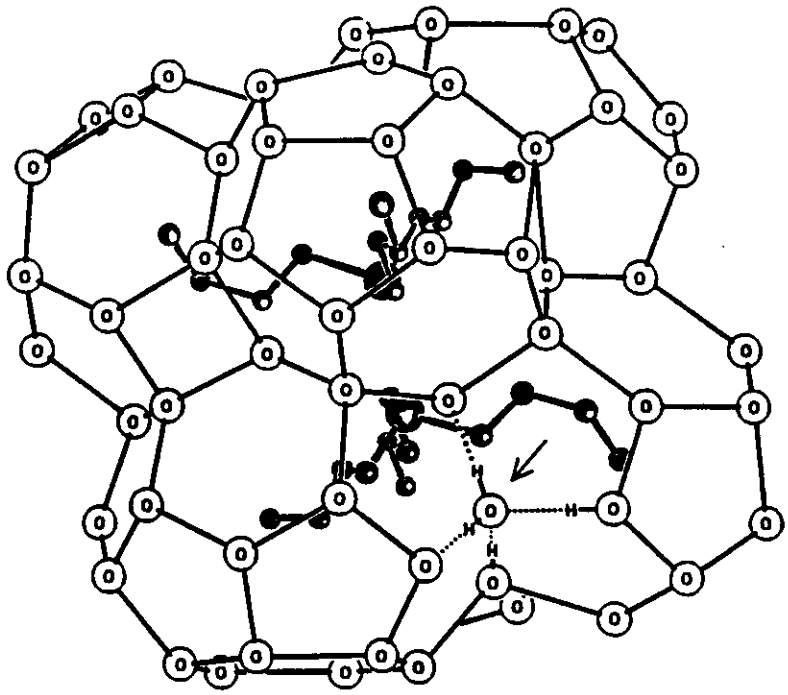
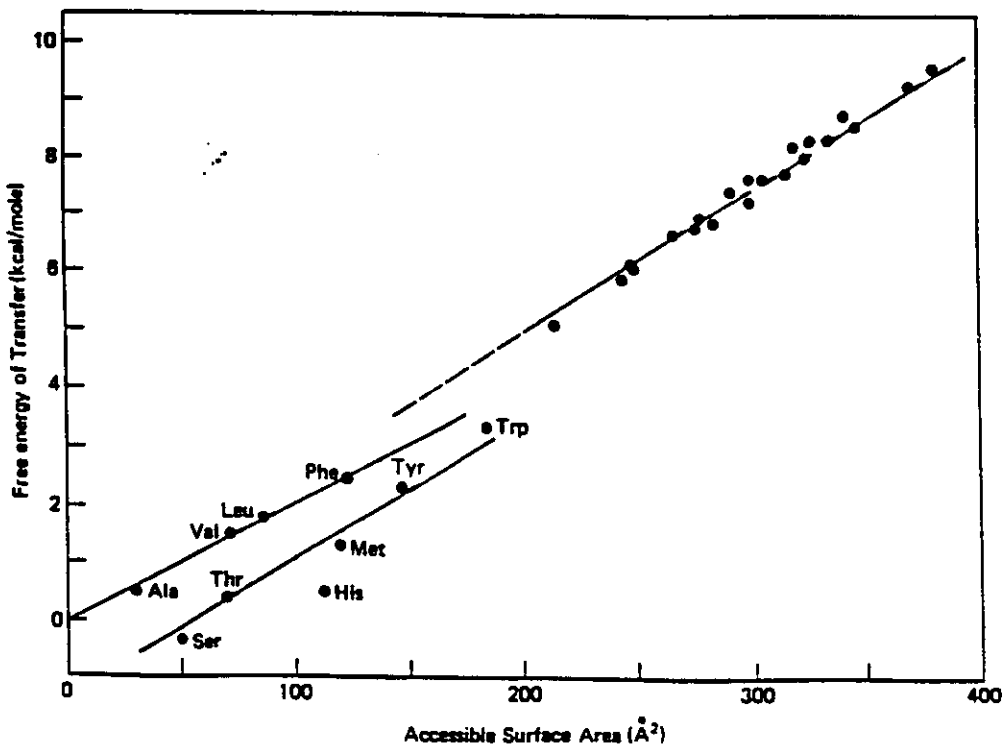


FIG. 2. Temperature dependence of partial heat capacity of lysozyme at pH values 2·0, 2·5 and 4·5.




Free energy of Transfer (kcal/mole)
Accessible Surface Area (Å²)
Trp
Phe
Tyr
Leu
Val
Met
Thr
His
Ala
Ser

Lecture 2
R. L. Baldwin
March 24, 1998

Peptide Helices as a Model for Protein Folding

Peptide helices exhibit the properties of protein folding at the micro-domain level. They are used to study two basic folding problems. (1) Are there autonomous folding units in the polypeptide backbone? Do protein fragments form helices whose structures match the helices in proteins? (2) Can the energetics of important folding interactions, such as the peptide H-bond or sidechain salt bridges be measured in peptide helices?

There is a snag, however. Helix formation is not a two-state reaction, even in short peptides. The ends of peptide helices are frayed, and helix formation begins with an unfavorable nucleation reaction (equilibrium constant 10^{-3}) and must then be pulled either by a favorable helix propagation parameter (greater than 1) or by specific sidechain interactions. In peptide helix studies, the helix propensity is defined as the helix propagation parameter of helix-coil theory: w in the Lifson-Roig theory, s in the Zimm-Bragg theory. It is a constant for a given residue, independent of position and neighboring residues. It became possible to use helix-coil theory to analyze peptide helix formation in water, without making assumptions about the values of sidechain interactions, only after it was discovered that alanine by itself can form a stable helix in water.

A. Early Studies of protein fragments

1. The first protein fragment found to show α-helix formation in water (the "C-peptide" of ribonuclease A, which contains the N-terminal α-helix, Thr 3 - His 12; J.E.Brown and W.A.Klee, Biochemistry 10, 470; 1971) proved to be stabilized by sidechain interactions present in the native protein (see review, R.L.Baldwin, Biophys. Chem. 55, 127; 1995). There is a salt bridge between Glu 2 and Arg 10 and a charge-aromatic interaction between protonated His 12 and Phe 8.

1. a. Thus, the C-peptide provides an example of an autonomous folding unit. Many later examples have been found. They provide strong support for the concept

that protein folding is hierarchic: i.e., that the folding process begins locally and local structures then interact to build up progressively more complex structures.

2. Most protein fragments from helix-containing regions show little helix formation in water, but can be induced to form helices by adding TFE (trifluoroethanol). The mechanism is probably strengthening of peptide H-bonds by TFE (P.Luo and R.L.Baldwin, Biochemistry 36, 8413; 1997).

3. Peptide helices, including the C-peptide helix, unfold with increasing temperature, like proteins. However, the mechanism of heat-induced unfolding, which is discussed below, is quite different for peptide helices and for proteins.

B. Alanine-based Peptide Helices

1. With the discovery of stable helix formation in water by alanine peptides (S.Marqusee et al., Proc. Natl. Acad. Sci. USA 86, 5286; 1989), it became possible to analyze peptide helix formation by helix-coil theory and to measure helix propensities as well as specific sidechain interactions.

1. a. By making a single Ala → Gly substitution at a series of positions in a 17-residue alanine-based peptide (A.Chakrabartty et al., Nature 351, 586; 1991), it was possible both to measure the ratio of the helix propensities for alanine/glycine (the current value is 35, C.A.Rohl et al., Protein Sci. 5, 2623; 1996) and to test the effect of fraying of the helix ends on making a helix-destabilizing mutation at different positions in a helix.

1. b. Although the earlier Zimm-Bragg theory (J. Chem. Phys. 31, 526; 1959) is better known, current studies of mixed sequence peptides use the Lifson-Roig theory (J. Chem. Phys. 34, 1963; 1961) because the helix propensity is assigned to the amino acid residue. In the Zimm-Bragg theory, which was developed for studying helix formation by homopolymers, the peptide group is the unit of helix formation rather than an amino acid residue.

2. The validity of helix-coil theory has been tested by using a series of repeating sequence peptides (AAKAA or EAAAKA repeat), with lengths varying from 6 to 51 residues, to measure the helix nucleation and propagation parameters. Two different methods of measuring these parameters have been used: fitting thermal unfolding curves monitored by circular dichroism (CD) at 222 nm (J.M.Scholtz et al., Biopolymers 31, 1463; 1991) and fitting amide proton exchange curves measured in conditions where the helical form is stable (C.A.Rohl et al., Biochemistry 31, 1263; 1992). Exchange in D_2O ($H \rightarrow D$) is irreversible and strongly dependent on chain length, because exchange occurs by fraying of the helix ends.

2. a. The two tests of helix-coil theory, and of these methods of measuring helix-coil parameters, are: (1) a single set of parameters should fit the results for peptides of varying chain lengths, and (2) the same values of the parameters should be obtained by both methods. These tests are satisfied by the data.

3. One residue at either end of the helix is half in, half out of the helix because only one of the two backbone (φ, ψ) angles is fixed. The original Zimm-Bragg and Lifson-Roig theories specify that these residues, termed N-cap and C-cap, should be coil rather than helix residues and changing the amino acid at these two positions should not affect helix stability.

3. a. Experiments show, however, that the choice of the N-cap residue does strongly affect helix-stability. The N-cap propensities of the various amino acids are not related to their helix propensities, and instead they are related to the frequencies of the various amino acids at the N-cap position in protein helices (A.Chakrabartty et al., Proc. Natl. Acad. Sci. USA 90, 11332; 1993).

4. The main point about the values of the helix propensities of the various amino acids is that there are large differences among them (the $\Delta\Delta G$ for Ala-Gly is 2 kcal/mol) and only alanine has a favorable helix propensity ($w > 1$). Current

values for alanine-based peptides are given by C.A.Rohl et al., Protein Sci. 5, 2623; 1996.

4. a. The rank order of helix propensities found in different systems, such as protein helices or dimeric coiled-coil peptides, is usually the same as in alanine-based peptides, but the differences among the various amino acids are larger in alanine-based peptides. The reason for this difference is not yet known.

C. Measurement of Sidechain and Other Interactions

1. Alanine, which has only a methyl group for a sidechain, forms the most stable helix. The interaction that drives helix formation is in the helix backbone itself: it is probably the peptide H-bond.

1. a. The enthalpy of forming the alanine helix has been measured calorimetrically for a 50 residue peptide. Its value is -1.1 ± 0.2 kcal/mol res (J.M.Scholtz et al., Proc. Natl. Acad. Sci. USA 88, 2584; 1991). This value agrees with the one obtained by fitting thermal unfolding curves to helix-coil theory.

1. b. The enthalpy change for forming the helix backbone, or peptide H-bond, may be different, however, when the helix is buried inside a protein, because H_2O that is H-bonded to peptide CO groups in a solvent-exposed helix is then stripped off (A.Ben Naim, J. Phys. Chem. 95, 1437; 1991).

1. c. The sidechains interfere with helix formation, and give lower values for the helix propensities of the amino acids, chiefly by steric clash with the helix backbone, which affects the distribution of sidechain rotamers and the sidechain entropy (T.P.Creamer and G.D.Rose, Proteins 19, 85; 1994).

2. Several types of specific sidechain interactions have been measured in alanine-based peptides: H-bonds, salt bridges, charge-aromatic interactions, nonpolar interactions and charge-helix dipole interactions. The measured value of an interaction needs to be corrected for the limited set of sidechain rotamers that can form the interaction. For example, only the trans, gauche + rotamers of

Glen, Asp (spaced i, I + 4) are able to form this sidechain H-bond interaction.

2. a. A summary of several sidechain interactions, corrected for rotamer choice is given by B.J.Stapley and A.J.Doig, J. Mol. Bio. 272, 465; 1997. Typical values are about -1 kcal/mol for a specific sidechain interaction.

2. b. It has been controversial whether salt bridge and sidechain H-bond interactions contribute to protein stability. Thus, it is noteworthy that these interactions can be measured straightforwardly and are demonstrated to be helix-stabilizing in peptide helices. A good example is the Gln, Asp (i, i + 4) sidechain H-bond, for which peptides containing 1, 2 or 3 pairs of interacting Gln, Asp residues show progressively higher helix contents (B.M.P.Huyghues-Despointes et al., Biochemistry 34, 13267; 1995). As a rule, sidechain interactions detected by frequency of occurrence in peptide helices are found to stabilize peptide helices, but the converse is not true. Thus, His, Asp (i, i + 3) and (i, i + 4) salt bridges stabilize peptide helices but do not occur at above-average frequency in protein helices (B.M.P.Huyghues-Despointes and R.L.Baldwin, Biochemistry 36, 1965; 1997).

3. By statistical fitting of all peptide helix data in the literature to helix-coil theory, L. Muñoz and L. Serrano (Nature Struct. Biol., 6, 399; 1994) were able to obtain helix propensities and sidechain interactions for a large number of interactions. On the whole, their values agree reasonably well with directly measured values from alanine-based peptides.

4. Values for nonpolar sidechain interactions have been estimated by Monte Carlo simulations by T. P. Creamer and G. D. Rose (Protein Sci. 4, 1305; 1995).

D. Helix Termination Signals

1. If protein folding is hierarchic and begins with the formation of secondary structure, then helix-termination signals should be encoded in the local sequence, rather than being determined by non-local interactions in the

tertiary structure.

1. a. A recent review of helix capping by R. Aurora and G. D. Rose (Protein Sci. 7, 21; 1998) surveys 1316 protein helices. The results show that termination signals are encoded locally and that 7 common motifs account for most of them. More than 80% of helices contain a hydrophobic capping motif, and about half that number contain a H-bonded capping motif.

1. b. There is a deficit of 4 mainchain peptide H bonds at each helix end, and making a sidechain-mainchain H-bond can stabilize a helix as well as terminate it.

1. c. NMR studies of peptides verify that these capping motifs can often be formed in peptide helices, where they help to stabilize the helix (P.C.Lyu et al., Biochemistry 32, 421; 1993).

Figures

Most figures are from two reviews: R. L. Baldwin (1995) Biophys. Chem. 55, 127 and A. Chakrabartty and R. L. Baldwin (1995) Adv. Protein Chem. 46, 141. Figures on page 4 are from B.M.P. Huyghues-Despointes et al. (1995) Biochemistry 34, 13267. The table of helix propensities in alanine-based peptides (page 3, upper) is from C. A. Rohl et al. (1996) Protein Sci. 5, 2623.

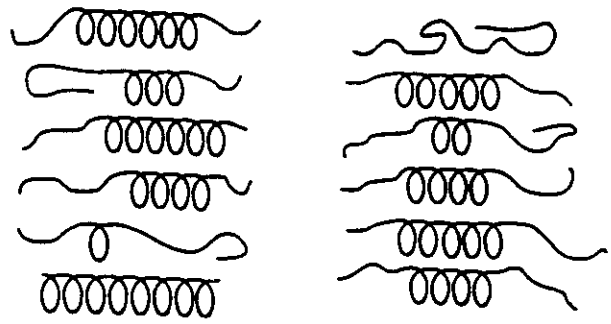


Figure 4 Diagram of frayed ends in partly helical peptide molecules. When the overall helix content of a peptide is measured to be 50% by circular dichroism, the system contains a distribution of molecules like those shown here.

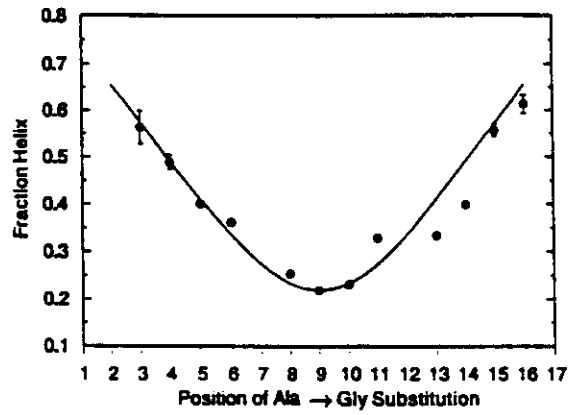


Figure 5 Helix content versus position of a single glycine residue in the reference peptide AcY-(KAAAA)$_3$-K(NH$_2$). Reprinted with permission from *Nature* (Chakrabartty *et al.*, 1991). Copyright 1991 Macmillan Magazines Limited.

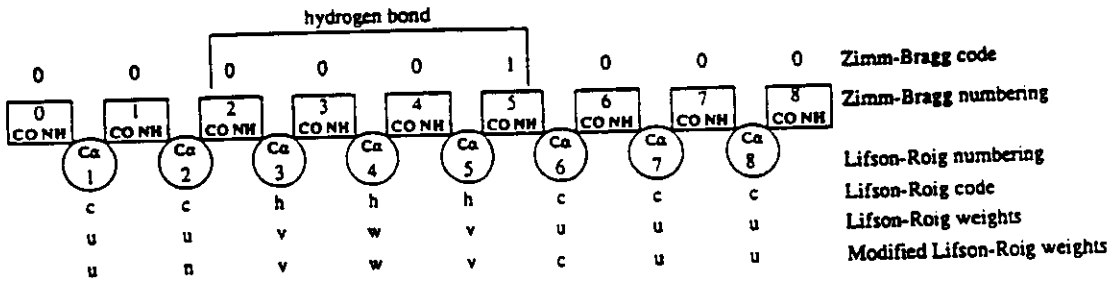


FIG. 2. Codes, numbering, and weighting of basic units of helix formation in the Zimm–Bragg (1959), Lifson–Roig (1961), and modified Lifson–Roig theories. Reprinted with permission from Doig *et al.*, 1994. Copyright (1994) American Chemical Society.

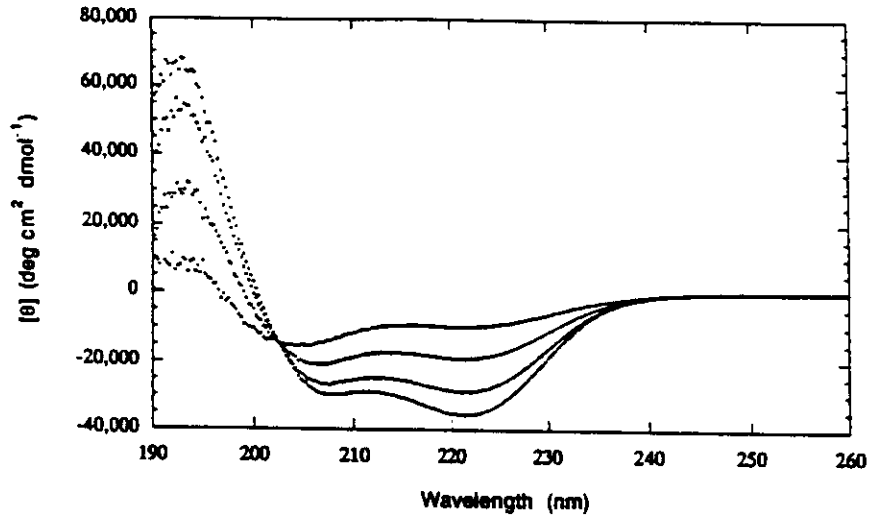


Figure 1 Helix unfolding monitored by CD spectra for a 50-residue peptide. Spectra of AcY-$(AEAAKA)_8$-F(NH_2) were recorded at various temperatures from 0°C (bottom curve at 222 nm) to 60°C (top curve at 222 nm) every 20°C. The peptide concentration was 8.5 μM in 1 mM potassium phosphate and 100 mM potassium fluoride at pH 7.00. The spectra were recorded on an Aviv 60DS spectropolarimeter in a 1.0-cm path length cuvette. Data are from Scholtz *et al.* (1991b).

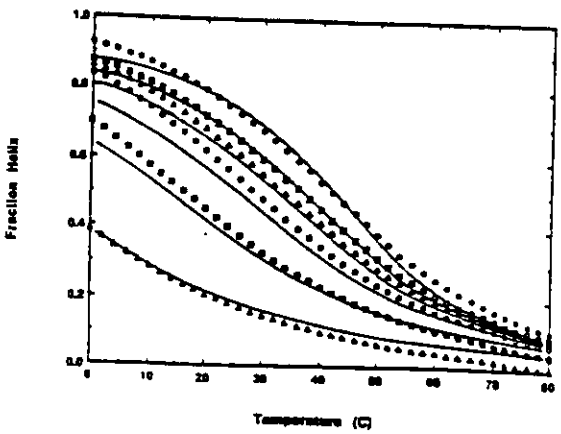


Fig. 2. Determining the parameters of the helix-coil transition by analyzing thermal unfolding curves monitored by CD for a series of peptides with chain lengths of 14 (bottom curve), 20, 26, 32, 38 and 50 (top curve) residues. The repeating sequence is AEAAKA, flanked by two marker residues: Ac–Tyr (N-terminal) and Phe–NH_2 (C-terminal). All unfolding curves are fitted by the same values of three adjustable parameters of helix-coil theory: the ...ciation constant, the enthalpy change per residue, and the ...ge value of the helix propagation parameter. Data from [25].

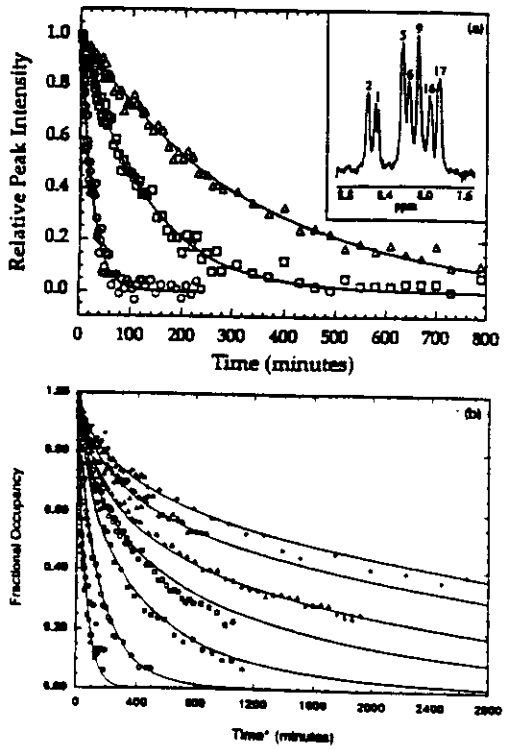


Table 2. *Helix propagation and N-cap propensities and free energies of amino acid residues in water at 273 K*

Residue	Helix propagation		N-cap	
	w	$\Delta G°$(helix)[a] (kcal/mol)	n	$\Delta\Delta G°$(N-cap)[b] (kcal/mol)
Ala	1.70	−0.27		
Glu^0	0.70	0.21		
Cys^0	0.32	0.64		
Cys^-			5.4	−0.92
Asp^-	0.38	0.54	6.6	−1.0
Glu^-	0.54	0.35	2.06	−0.39
Phe	0.27	0.73	2.06	−0.39
Gly	0.048	1.7	3.9	−0.74
His^+	0.22	0.84		
Ile	0.46	0.44	1.57	−0.25
His^0	0.36	0.57	2.12	−0.41
Lys^+	1.00	0.019	0.72	0.18
Leu	0.87	0.095	2.06	−0.39
Met	0.65	0.25	1.31	−0.15
Asn	0.29	0.69	6.8	−1.0
Asp^0	0.40	0.52		
Pro	<0.001	>3.8	1.35	−0.16
Gln	0.62	0.28	0.12	1.2
Arg^+	1.14	−0.052	1.00	0.0
Ser	0.40	0.52	3.9	−0.74
Thr	0.18	0.95	2.23	−0.44
Val	0.25	0.77	0.96	0.022
Trp	0.29	0.69	3.6	−0.70
Tyr	0.48	0.42	4.9	−0.96
Acetyl			5.9	−0.86

[a] $\Delta G(\text{helix}) = -RT \ln[w/(1 + v)]$, where $v = 0.036$.
[b] $\Delta\Delta G(\text{N-cap}) = -RT \ln[n]$. N-cap propensities and free energies are relative to Ala.

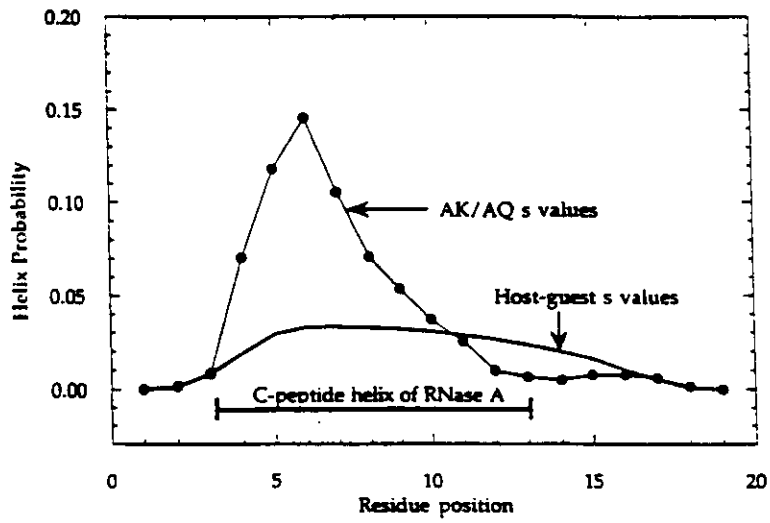


FIG. 5. Calculated distribution of helicity in the S peptide of RNase A. The calculation used the Lifson–Roig theory modified to include helix capping with either $\sigma = 0.003$, n and s values from alanine-based peptides (Chakrabartty *et al.*, 1994), or $\sigma = 0.0001$ and s values from host–guest studies (Wojcik *et al.*, 1990). The bar indicates location of the C-peptide helix in the crystal structure of RNase A.

A

Peptide	Sequence
1QD	Ac-AAQAAADQAAAAQAAY-NH_2
1QDref	Ac-AAQAADAQAAAAQAAY-NH_2
2QD	Ac-AAQAAADQAAADAAAGY-NH_2
2QDref	Ac-AAAQAADAQAADAAAGY-NH_2
3QD	Ac-AAQAAADQAAADQAAADAAAGY-NH_2
3QDref	Ac-AAAQAADAQAADAQAADAAAGY-NH_2

B

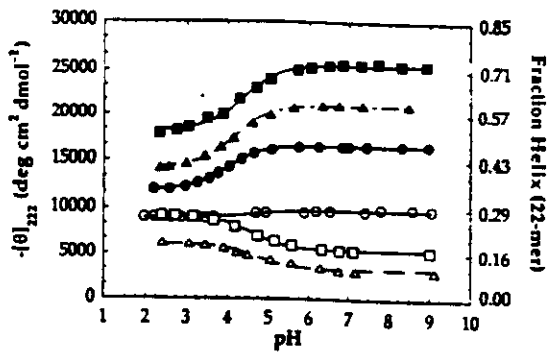


FIGURE 2: Measurement by circular dichroism of the increase in helix content caused by the Gln·Asp (i, $i+4$) interaction in peptides with 1, 2, or 3 interacting pairs of Gln, Asp residues. A. Sequences of the peptides studied and of the corresponding reference peptides. B. pH titration curves of helix content measured by circular dichroism. Filled symbols show the test peptides, open symbols show the corresponding reference peptides: (●) 1QD. (▲) 2QD. (■) 3QD. $[\Theta]_{222}$ is the mean residue ellipticity measured at 222 nm.

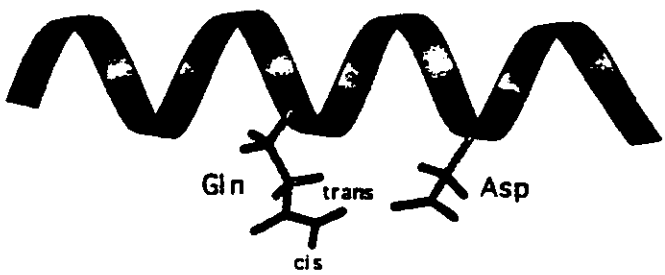


Table 2. Frequencies of (i, i + 4) Side-Chain H-Bond Interactions in Protein Helices

sequence	no.[a]	H-bond[b]	i *trans*[c]	$i+4$ *gauche*+[d]
QD	35 (18.5)	19	29 (14.4)	32 (27.6)
QE	52 (31.7)	18	32 (21.4)	25 (31.8)
EQ	29 (31.8)	2	11 (11.4)	18 (18.0)
DQ	22 (21.3)	1	5 (3.9)	13 (13.6)

Lecture #3
R. L. Baldwin
March 25, 1998

The Study of Protein Folding Intermediates and Folding Pathways

The problem is to understand how the amino acid sequence of a protein determines its 3D structure by analyzing stages in the folding process. What are the structures of the folding intermediates and what are the interactions that stabilize them? Are the intermediates and pathway unique, or are there competing pathways for folding and unfolding?

To simplify the problem, our focus here is on the folding reactions of small, single-domain proteins. A current puzzle concerns two classes of folding reactions: one class of small proteins shows readily observable folding intermediates while a second class folds at least as rapidly without detectable intermediates.

A. Tests for Folding Intermediates

1. In native conditions, the heat-induced or denaturant-induced unfolding transitions of most small proteins show no detectable intermediates: they are two-state ($U \rightleftharpoons N$) folding reactions.

1. a. Nevertheless, their folding kinetics often reveal transiently populated intermediates (see below).

1. b. A few proteins, such as α-lactalbumin (α-LA), do show an equilibrium intermediate. A standard test for the presence of an intermediate is to compare the unfolding transition curves measured by probes of tertiary structure (such as Trp fluorescence) and of secondary structure (such as far-UV CD): see K.Kuwajima et al., J. Mol. Biol. 1-6, 359; 1976. The tertiary structure unfolds before the secondary structure (i.e., at a lower temperature or denaturant concentration) when an intermediate is present.

1. c. In non-native conditions, such as acid pH, folding intermediates are found more often. The native structure becomes unstable at acid pH while the folding intermediate remains more stable than the acid-unfolded form, particularly in the

presence of a stabilizing anion (only a few mM is needed) such as sulfate or perchlorate (Y.Goto et al., Biochemistry 29, 3480; 1990).

1. d. These equilibrium intermediates found at acid pH are often called molten globule intermediates because they appear to have no fixed tertiary structure while retaining some or all of the secondary structure of the native protein: see below.

2. Two standard tests for kinetic folding intermediate are as follows.

2. a. (1) The folding reaction is two-state (no apparent intermediates) if it follows single-exponential kinetics both in the folding and unfolding directions, and if the entire amplitude predicted from the equilibrium transition curve is found in the kinetic curve, meaning that there are no additional faster or slower steps in the reaction. The ratio of the rate constants for folding and unfolding should give the equilibrium constant.

2. b. (2) The folding or unfolding kinetics should be the same, if there are no intermediates, when measured by probes of secondary and tertiary structure.

2. c. Experience shows that folding kinetics provide a much more sensitive test for intermediates than equilibrium unfolding results, at least in native conditions. Consequently, attention has been focused until recently on kinetic folding intermediates. At least three proteins are now known, however, in which a major kinetic intermediate formed in native conditions corresponds closely to an equilibrium intermediate formed at acid pH: α-lactalbumin, apomyoglobin, and RNase H. Some results are discussed below. Because of the obvious advantages of determining structure and properties for equilibrium intermediates, these protein systems are now being studied intensively.

B. Characterization of the Transition State for a Two-State Folding Reaction

1. Typically lnk depends linearly on denaturant concentration. (c), where k is the rate constant either for folding (k_f) or unfolding (k_u).

$$\ln k_f = \ln k_f(H_2O) + (m_f/RT)\ C \qquad (1a)$$

$$\ln k_u = \ln k_u(H_2O) + (m_u/RT)\ C \qquad (1b)$$

The equilibrium m-value equals $m_f - m_u$. For folding, lnk decreases with C while for unfolding lnk increases. This behavior gives rise to a V-shaped or "chevron" plot with a minimum rate at the midpoint of the unfolding transition.

1. a. The ratio m_f/m is a measure of the surface area buried during folding to the transition state divided by the surface area buried in the complete folding reaction (C.Tanford, Adv. Protein Chem. 24, 2; 1970). This interpretation assumes that the number of sites available for interaction with a denaturant is simply proportional to exposed surface area.

2. Mutational analysis of the transition state. The transition state for folding is assumed to be in equilibrium with the unfolded, or denatured, state:

$$\ln k_f = \ln k_o - (\Delta G^{\ddagger}/RT) \qquad (2)$$

$\Delta G^{\ddagger}$ is the difference in free energy between the unfolded state and transition state, and k_o is an unknown "pre-factor" rate constant, which is assumed to be constant, not affected by mutation. Thus, although $\Delta G^{\ddagger}$ itself is not known because k_o is unknown, the changes in $\Delta G^{\ddagger}$ for a series of mutations can be measured.

2. a. In the "Brønsted plot", $\Delta\Delta G^{\ddagger}$ is plotted against $\Delta\Delta G$ for a series of mutants, where $\Delta\Delta G^{\ddagger}$ is the difference in $\Delta G^{\ddagger}$ between wild type and a mutant, and $\Delta\Delta G$ is the corresponding difference between the free energy of folding. For a two-state folding reaction, a straight line is often obtained: see data for CI2 given by L.S.Itzhaki et al., J. Mol. Biol. 254, 260; 1995. The slope (0.7 for CI2) gives the relative effect of a mutation on the stability of the transition state divided by its effect on the native state.

2. b. The φ-value is a closely related quantity given for a single mutation: φ is $\Delta\Delta G^{\ddagger}/\Delta\Delta G$, where the difference between wild type and mutant is taken. Often $\Delta\Delta G^{\ddagger}$ is found from the unfolding rates of mutant and wild type, rather than from the

folding rates. In considering values of φ, it is necessary to check whether they are based on folding or unfolding rates.

C. Characterize Folding Intermediates

1. Structural characterization.

1. a. The main tool used thus far is NMR-amide proton exchange (see F.M.Hughson et al., Science 249, 1544; 1990). When used to characterize an equilibrium intermediate, it is diluted from H_2O into D_2O, samples are taken at a series of times and exchange is quenched (often by refolding rapidly to the native form) and the 2D ^{1}H-NMR spectrum of the folded protein is measured to determine the extent of exchange of every peptide NH proton in the spectrum, for each time point. The results give the locations of the protected NH protons and their protection factors (PF). The PF is defined as the exchange rate of a NH proton in the unfolded form divided by its exchange rate in the intermediate.

1. b. Folding intermediates often have PF values from 10 to 10^3. Native proteins have a core of protected protons with PF values near 10^8.

1. c. For a kinetic folding intermediate, the same technique is used with a stopped-flow pulse-and-quench apparatus. Exchange is base catalyzed and occurs in about 1 millisecond at pH9, 10°C, for an unprotected proton. Thus, the folding intermediate is exposed to a pH near 9 for only a few msec.

1. d. Quite recently, it has been possible to assign the backbone resonance lines of an equilibrium folding intermediate of apomyoglobin (E.Eliezer et al., Nature Struct. Biol. 5, 148; 1998). This opens the way to detailed structural characterization.

1. e. Typically sidechain proton resonances are unresolved in spectra of equilibrium folding intermediates, perhaps because of exchange broadening, but they may become resolved (using ^{15}N-labeled protein) in the spectrum of the unfolded protein. A technique based on this observation has been used to monitor the urea unfolding of different regions of an α-LA folding intermediate: B.A.Schulman

et al., Nature Struct. Biol. 4, 630; 1997.

2. Measurement of stability and stabilizing interactions.

2. a. Typically ΔG (unfolding) can be measured only for two-state unfolding reactions without intermediates. Thus, the first problem has been to determine if equilibrium folding intermediates show two-state unfolding reactions.

2. b. The pH4 folding intermediate of apomyoglobin does show two-state urea-induced unfolding in specified conditions (pH 4.2, 4°C, 4 mM Na citrate). The unfolding curves measured by probes of secondary structure (far-UV CD) and tertiary structure (Trp fluorescence) are superimposable (M.S.Kay and R.L.Baldwin, Nature Struct. Biol. 3, 439; 1996) and the unfolding and refolding kinetics measured by Trp fluorescence are two-state, with the possible exception of a burst phase in unfolding (M.Jamin and R.L.Baldwin, Nature Struct. Biol. 3, 613; 1996).

2. c. When helix-destabilizing mutations are made (substitution of Gly or Pro at a solvent-exposed position), the results confirm that wild type folding is strongly cooperative, but the cooperativity becomes weaker as the stability is reduced (Y.Luo et al., Nature Struct. Biol. 4, 925; 1997).

2. d. Some tertiary interactions that are present in the native protein may contribute to the stability of the intermediate: hydrophobic packing interactions (Kay and Baldwin, reference above; also in cytochrome c ((J.L.Marmorino and G.J.Pielak, Biochemistry 34, 3140; 1995) and a tertiary salt bridge in barnase (M.Oliveberg and A.R.Fersht, Biochemistry 35, 6795; 1996).

2. e. On the other hand, the α-lactalbumin folding intermediate folds non-cooperatively, as judged both by proline mutagenesis (B.A.Schulman et al., Nature Struct. Biol. 4, 630; 1997) and by monitoring urea-induced unfolding with 2D NMR (B.A. Schulman et al., Nature Struct. Biol. 4, 630; 1997).

2. f. An engineered version of α-lactalbumin that lacks the β-sheet domain and 2 of the 4 disulfide bonds has the correct tertiary fold of the native protein, although it has

a molten globule conformation, according to the pattern of reforming the two remaining disulfide bonds (Z.Peng and P.S.Kim, Biochemistry 33, 2136; 1994).

Figures

1. Upper: GdmCl-induced unfolding of α-lactalbumin measured by near-UV and far-UV CD (Kuwajima et al, 1976).

   Lower: anion stabilization of the molten globule intermediate of cytochome c at pH 2 (Goto et al., 1990).

2. Left: Brønsted plot for CI 2 (Itzhaki et al., 1995).

   Right: Effects of GdmCl (upper) and temperature (lower) on the kinetics of the unfolding transition of protein L: from M. Scalley and D. Baker (1997) Proc. Natl. Acad. Sci. USA 94, 10636.

3. Upper: table of protection factors measured for native apomyoglobin and for the pH 4 intermediate by Hughson et al. (1990).

   Lower: stabilization of a second form of the pH 4 intermediate of apoMb by 20 mM trichloroacetate: from S. N. Loh et al. (1995) Proc. Natl. Acad. Sci. USA 92, 5446.

4. Unfolding curves and stability measurements on the pH 4 intermediate of apoMb: from Kay & Baldwin (1996).

5. NMR characterization of the pH 4 intermediate of apoMb: from D. Eliezer et al. (1998).

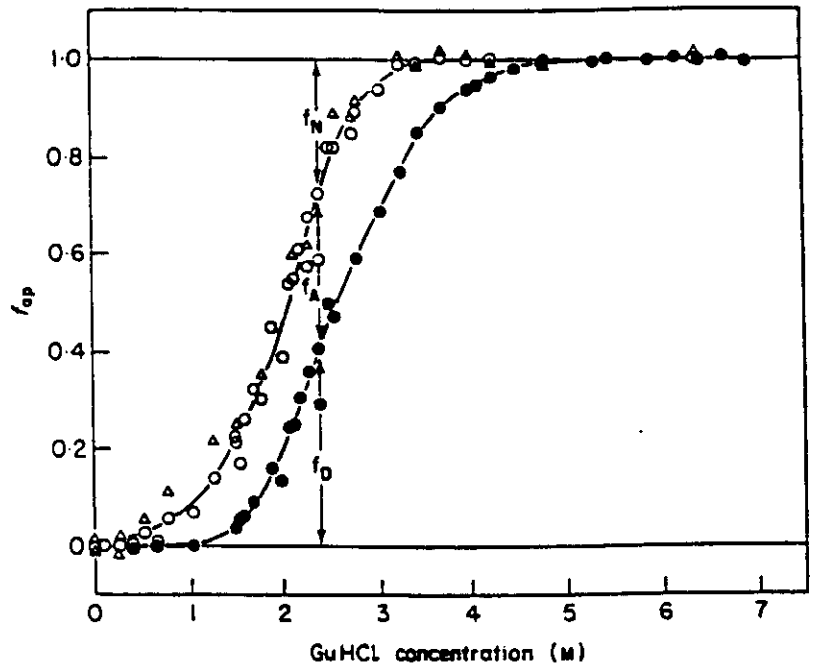


FIG. 2. The dependence of f_{ap} on GuHCl concentration (pH 6·65); f_{ap} is calculated from the molar ellipticity at 270 nm (○), at 296 nm (△) and at 222 nm (●).

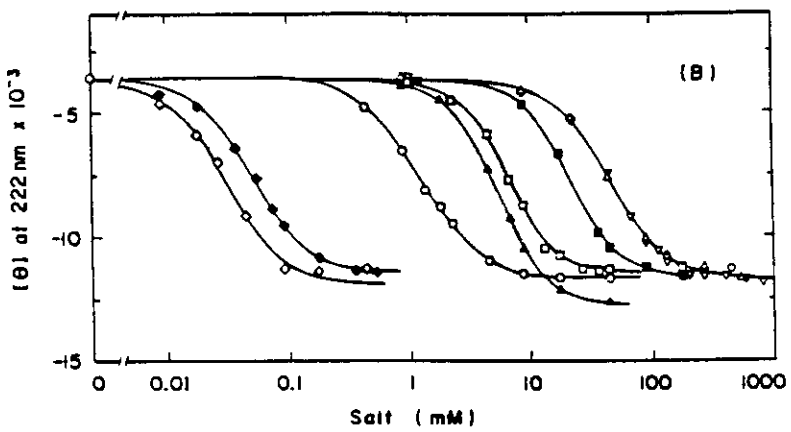
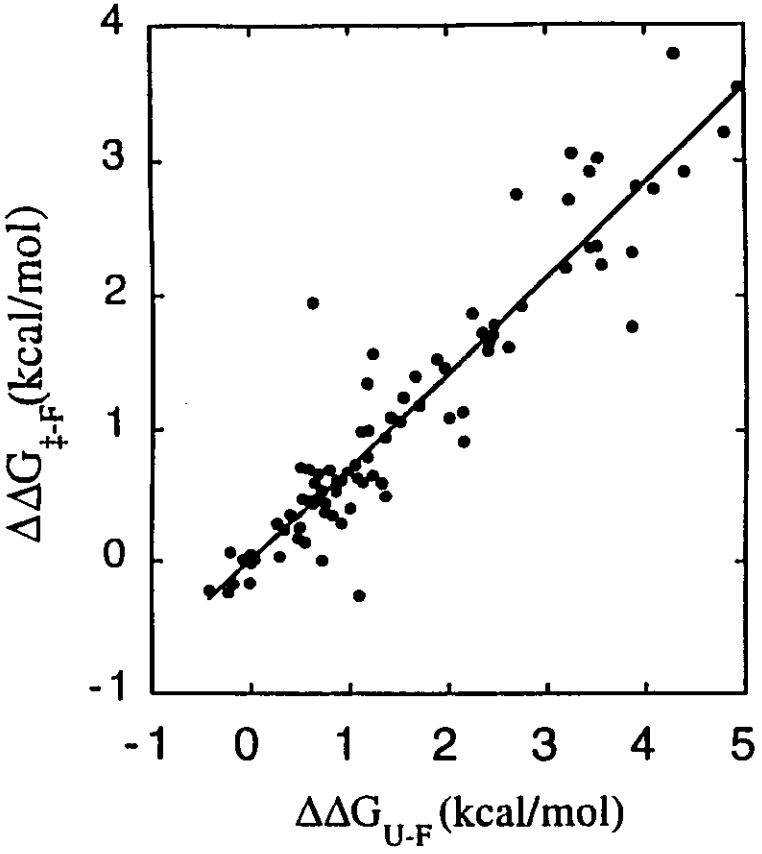


FIGURE 3: Acid-induced (A) or salt-induced (B) conformational transitions of cytochrome *c* in the presence of 18 mM HCl measured by the ellipticity at 222 nm at 20 °C. Acids used were HCl (△), TFAH (■), perchloric acid (□), TCAH (▲), and sulfuric acid (○). Salts were NaCl (△), KCl (▽), TFANa (■), sodium perchlorate (□), TCANa (▲), sodium sulfate (○), potassium ferrocyanide (◆), and potassium ferricyanide (◇).



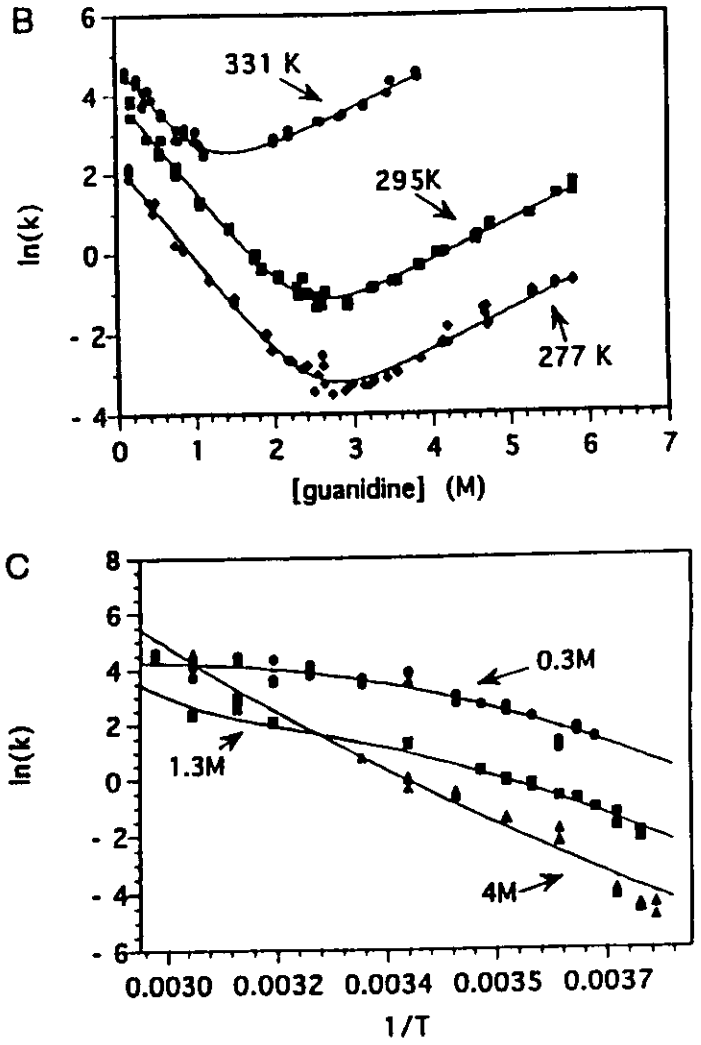


FIG. 1. Temperature and denaturant dependence of protein L folding kinetics. (*A*) Folding rate versus temperature and denaturant concentration.

Table 1. Amide proton protection factors. Protection factors (k_{exp}/k_{obs}) are given for amide protons for which assignments have been made and for which exchange is effectively quenched by reconstitution. Time points for exchange-out were collected as described in the legend to Fig. 3. The extent of exchange at each time point was determined by calculating the volume integral of each NH–C_αH cross-peak in the COSY spectrum. The data at each time point were normalized with reference to cross-peaks between nonexchangeable protons. Normalized cross-peak intensities for duplicate experiments differed on average by <25 percent. Five time points with exchange-out intervals ranging from <1 second to 3000 minutes were collected for native apo-Mb; seven time points with exchange-out intervals ranging from <1 minute to 630 minutes were collected for apo-Mb in the I (intermediate) form. Nonlinear least squares fitting to an exponentially decaying function were used to determine the observed exchange rates (k_{obs}). Solvent-exposed exchange rates (k_{exp}) were calculated according to (*10*). ND, Not determined. For pH 4.2 data, Leu^{40}, Gly^{45}, and Gly^{73} were lost during reconstitution. For pH 6 data, Leu^{76} could not be measured due to overlap with Lys^{34}.

Residue	Helix position (holo-Mb)	*P* (protection factor) Native (pH 6.0) (apo-Mb)	*P* (protection factor) I (pH 4.2) (apo-Mb)	Residue	Helix position (holo-Mb)	*P* (protection factor) Native (pH 6.0) (apo-Mb)	*P* (protection factor) I (pH 4.2) (apo-Mb)	Residue	Helix position (holo-Mb)	*P* (protection factor) Native (pH 6.0) (apo-Mb)	*P* (protection factor) I (pH 4.2) (apo-Mb)
Leu^{9}	A7	9,000	90	Phe^{43}	CD1	30	0.6	Lys^{102}	G3	<10	5
Val^{10}	A8	100,000	200	Gly^{45}	E8	2,000	ND	Tyr^{103}	G4	<10	1
Leu^{11}	A9	>100,000	60	Val^{68}	E9	3,000	1	Leu^{104}	G5	10	8
Trp^{14}	A12	>100,000	20	Val^{68}	E11	1,000	0.4	Ile^{107}	G8	2,000	40
Val^{17}	A15	80,000	40	Leu^{69}	E12	5,000	1	Ala^{110}	G11	6,000	7
Glu^{18}	A16	9,000	10	Thr^{70}	E13	4,000	2	Ile^{112}	G13	60,000	80
Ile^{28}	B9	4,000	2	Ala^{71}	E14	30	0.3	Val^{114}	G15	200,000	30
Leu^{29}	B10	20,000	3	Leu^{72}	E15	100	0.4	Leu^{115}	G16	60,000	7
Ile^{30}	B11	70,000	10	Gly^{73}	E16	3,000	ND	Phe^{123}	GH5	1,000	<1
Arg^{31}	B12	5,000	2	Ala^{74}	E17	1,000	0.8	Lys^{133}	H10	6,000	5
Phe^{33}	B14	6,000	5	Ile^{75}	E18	2,000	1	Phe^{138}	H15	7,000	10
Lys^{34}	B15	4,000	0.9	Leu^{76}	E19	ND	0.6	Ile^{142}	H19	200	60
Leu^{40}	C5	20,000	ND	Lys^{77}	E20	500	<1	Ala^{143}	H20	20	20

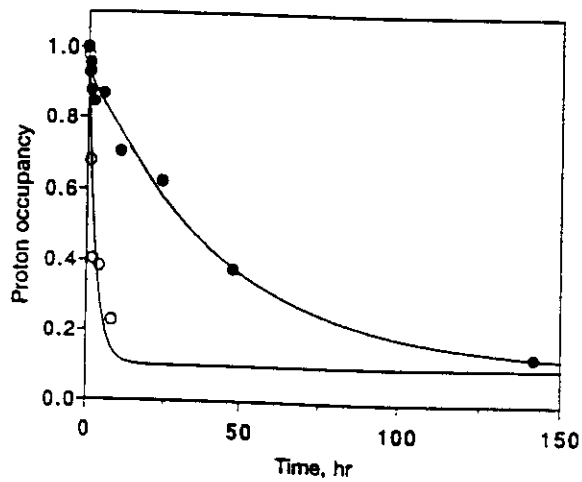


FIG. 4. Representative plot of proton occupancy as a function of exchange time for the B-helix residue Ile-30 in the I_1 state (no CCl_3COONa present) (○) and in the I_2 state (20 mM CCl_3COONa present) (●). Lines are the best fit of the data to single exponential decays.

Table 1 Apparent stabilities of mutants (kcal mol^{-1})

Mutant	Location	Intermediate $-\Delta G_{IU}$[1] ($\Delta\Delta G_{wt}$)	Native $-\Delta G_{NU}$[2] ($\Delta\Delta G_{wt}$)
wild type		2.9	5.6
W7F	A/H interface	1.8 (1.1)	4.7 (0.9)
W14F	A/E interface	2.4 (0.5)	4.5 (1.1)
M131A	A/G/H interface	2.0 (0.9)	3.4 (2.2)
F123K	GH loop	2.2 (0.7)	3.5 (2.1)
A130K	A/H interface	2.0 (0.9)	3.5 (2.1)
A130L	A/H interface	3.0[3] (-0.1)	4.7 (0.9)
H36Q	BC/G interface	2.8 (0.1)	4.8 (0.8)
V68T	E helix	2.7 (0.2)	5.0 (0.6)

[1]Stability of the intermediate to urea-induced unfolding at pH 4.2, 4 °C, 4 mM citrate; the change from wild type is given in parentheses.
[2]Stability of native apoMb to urea-induced unfolding at pH 7.8, 4 °C, 10 mM Hepes; the change from wild type is given in parentheses.
[3]This mutant, in contrast to the other mutants in this table gives slightly different results when the transition curve is measured either by CD or by fluorescence (see Fig. 7). The average of the fluorescence (2.9 kcal mol^{-1}) and CD (3.1 kcal mol^{-1}) results is given here.

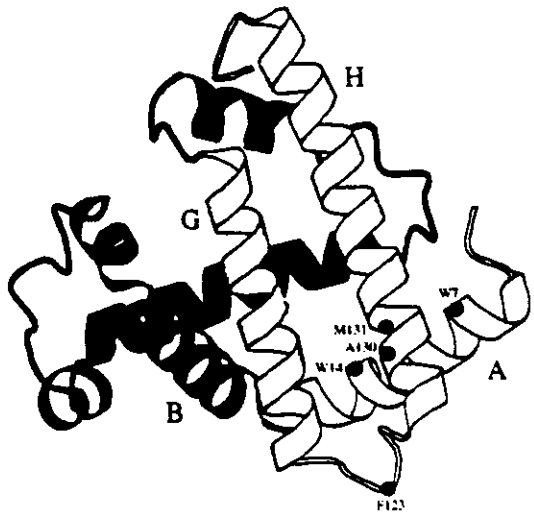


Fig. 1 Molscript[39] diagram of sperm whale myoglobin with AGH subdomain highlighted (white). Mutated residues within the AGH subdomain are highlighted.

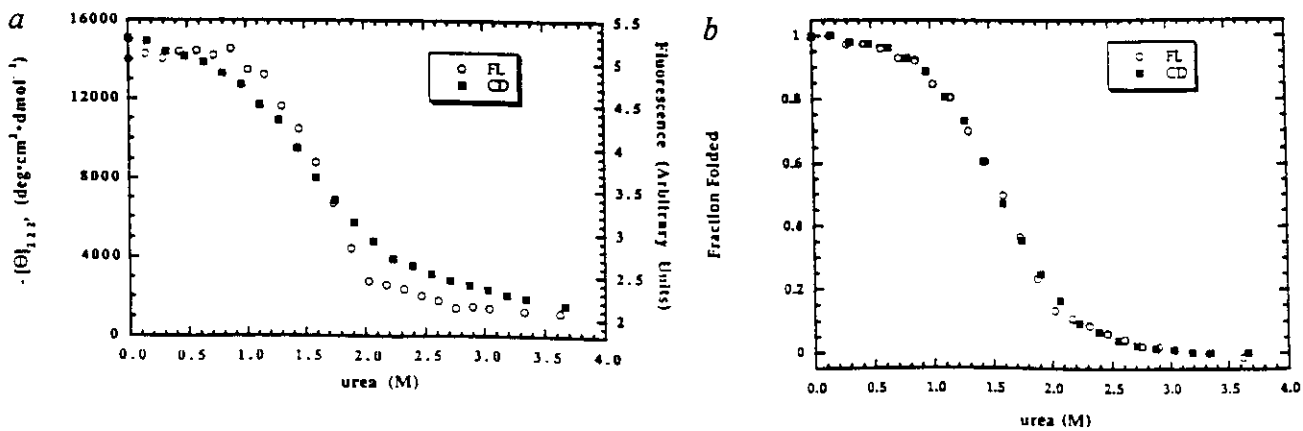


Fig. 2 CD and fluorescence data for urea-induced unfolding of wild-type apoMb intermediate (pH 4.2, 4 °C, 4 mM citrate): *a*, raw data and *b*, normalized unfolding curves.

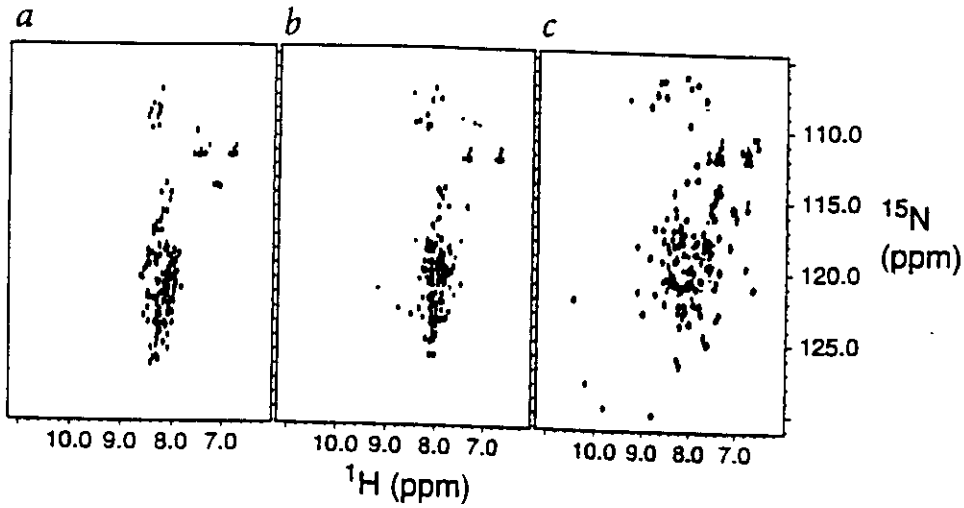


Fig. 1 750 MHz ^{1}H-^{15}N heteronuclear single quantum correlated NMR spectra of apomyoglobin in *a*, the acid-unfolded state (pH 2.3, 25 °C), *b*, the molten globule state (pH 4.1, 50 °C) and *c*, the native state (pH 6.1, 35 °C).

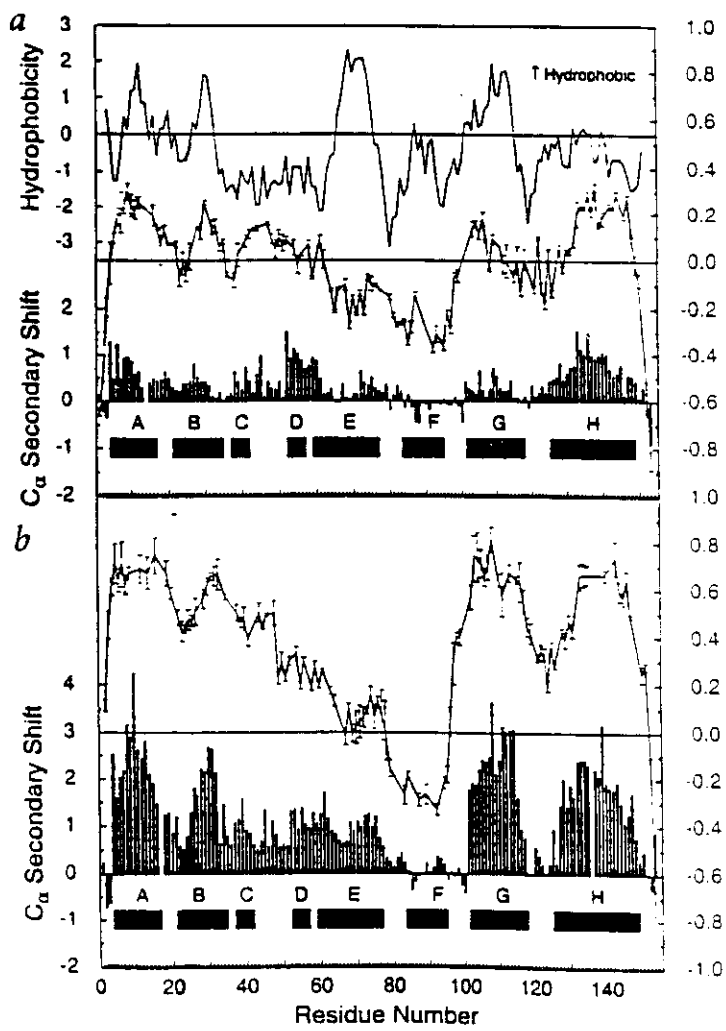


Fig. 2 Summary of secondary $^{13}C\alpha$ chemical shifts and (^{1}H)-^{15}N heteronuclear NOEs for apomyoglobin in the *a*, pH 2.3 state and *b*, pH 4.1 molten globule state. Black histogram and axis: deviation of the $^{13}C\alpha$ chemical shift from random coil values (p.p.m.). Red data and axis: magnitude of the (^{1}H)-^{15}N heteronuclear NOE for the backbone amide nitrogen atoms. Experimental uncertainties are indicated by vertical bars. Green curve and axis: hydrophobicity estimated by published methods[34], using a seven-residue sampling window. Positive values indicate hydrophobic regions and negative values indicate hydrophilic regions of the amino acid sequence. The horizontal black bars indicate the location of the helices in the X-ray structure of holomyoglobin (PDB accession number 2mbw)[57]. Random coil $^{13}C\alpha$ chemical shifts were obtained from Wishart *et al.*[29], with appropriate corrections for residues preceding proline. Chemical shifts of Glu and Asp resonances in the pH 2.3 state were corrected for the effect of pH.