doi:10.1016/j.jmb.2007.05.078 J. Mol. Biol. (2007) 371, 283–301

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

REVIEW

Energetics of Protein Folding

Robert L. Baldwin

Department of Biochemistry Beckman Center, Stanford University Medical Center Stanford, CA 94305, USA

The energetics of protein folding determine the 3D structure of a folded protein. Knowledge of the energetics is needed to predict the 3D structure from the amino acid sequence or to modify the structure by protein engineering. Recent developments are discussed: major factors are reviewed and auxiliary factors are discussed briefly. Major factors include the hydrophobic factor (burial of non-polar surface area) and van der Waals interactions together with peptide hydrogen bonds and peptide solvation. The long-standing model for the hydrophobic factor (free energy change proportional to buried non-polar surface area) is contrasted with the packing–desolvation model and the approximate nature of the proportionality between free energy and apolar surface area is discussed. Recent energetic studies of forming peptide hydrogen bonds (gas phase) are reviewed together with studies of peptide solvation in solution. Closer agreement is achieved between the 1995 values for protein unfolding enthalpies in vacuum given by Lazaridis-Archontis-Karplus and Makhatadze-Privalov when the solvation enthalpy of the peptide group is taken from electrostatic calculations. Auxiliary factors in folding energetics include salt bridges and side-chain hydrogen bonds, disulfide bridges, and propensities to form α-helices and β-structure. Backbone conformational entropy is a major energetic factor which is discussed only briefly for lack of knowledge.

Keywords: peptide hydrogen bonds; peptide solvation; buried non-polar surface; van der Waals interactions; backbone entropy

Introduction

Protein folding is acknowledged today to be a self-assembly process, guided by finding the structure of minimum free energy. Thus, predicting *de novo* the 3D structure of a protein from its amino acid sequence requires knowledge of the folding energetics. Likewise, understanding the structure of a complex formed between two proteins, or between a protein and a specific ligand, requires knowing the folding energetics, which also governs the dynamics of protein conformational change. If a process is not energetically feasible, it does not happen. Knowledge of protein folding energetics is needed to understand such diverse cellular processes as sending instructions along signaling pathways and making antibodies against specific antigens.

Proteins are synthesized as unfolded polypeptide chains and they fold after synthesis in order to become active. Anfinsen[1] realized that the driving force for folding is the gradient of free energy and the search for the free energy minimum gives the 3D structure, which is the most stable structure. Most methods of predicting structure focus, however, on finding the structure that has minimum enthalpy

Abbreviations used: H-bond, hydrogen bond; ASA, water-accessible surface area; ΔG_{hyd}, free energy change for burial of non-polar surface; k_h, proportionality factor between ΔG_{hyd} and ASA; ΔG_c, free energy of cavity formation; U, N, unfolded and native forms, respectively; CED, cohesive energy density; E_a, van der Waals interaction energy; γ, surface tension; ϕ, ψ, peptide backbone angles; ESF, electrostatic solvation free energy; P_{II}, polyproline II; ΔS_{res}, per residue change in backbone entropy; BPTI, bovine pancreatic trypsin inhibitor; T_m, temperature midpoint of a thermal unfolding transition; MP, Makhatadze and Privalov; LAK, Lazaridis, Archontis and Karplus; DFT, density functional theory.

E-mail address of the corresponding author: baldwinb@stanford.edu

rather than minimum free energy. The reason is that the enthalpy can be determined more reliably than the free energy and the minimum enthalpy structure is likely to give also the minimum free energy. This review follows the same logic and focuses chiefly on the enthalpy of folding, except for the discussion of the hydrophobic factor, where entropy and enthalpy are intertwined. The change in backbone entropy upon folding is reviewed briefly. Because proteins begin their biological lives as unfolded chains, it is always necessary to consider the equilibrium between the native and unfolded forms when considering whether a given factor, such as the peptide hydrogen bond, stabilizes folding.

The energetic factors that control protein folding are the subject of a classic review by Dill[2] in 1990 and the outlines of the subject remain much the same today. Dissection of protein unfolding enthalpies into factors such as hydrogen bonds and van der Waals energies was the subject of two major reviews in 1995: experimental work and empirical analysis of the results, reviewed by Makhatadze and Privalov,[3] and theoretical analysis by Lazaridis *et al.*[4] Serious disagreement was found[3,4] concerning the sizes in vacuum of the unfolding enthalpy and its individual components, especially the hydrogen bond enthalpy. A possible resolution of the controversy is discussed here.

The present review focuses on two major energetic factors, the hydrophobic factor (burial of non-polar surface area) and peptide hydrogen bonds (H-bonds), whose energetics are connected with solvation of the peptide group. The hydrophobic factor has long been regarded as the most important single factor and it is the most studied. It is difficult to separate the hydrophobic factor from van der Waals (or packing, or dispersion force) interactions and the two are discussed together here. Much has been written about the physical chemistry of the hydrophobic factor and how to obtain its contribution to the energetics of protein folding, and this subject is discussed first. The energetics of making peptide H-bonds is only beginning to be understood quantitatively. Work on this problem has been deferred until recently, partly for lack of viable approaches, partly because most estimates suggest that peptide H-bonds have only marginal stability. However, there are clear indications today that peptide H-bonds are a major factor in protein folding energetics and work on this problem is now accelerating.

There are also auxiliary factors in protein folding energetics, including electrostatic interactions among ionized side-chains, α-helix and β-sheet propensities (which are closely connected with the energetics of peptide H-bonds), and specific pairwise interactions among side-chains, such as H-bonds and salt bridges. The status of these factors is described briefly. The change in backbone conformational entropy upon folding is a subject of major energetic importance, but little is known. Its status is described briefly together with the auxiliary factors.

The hydrophobic factor in protein folding and Kauzmann's model

Hydrocarbon and other non-polar molecules have low solubilities in water but good solubilities in non-aqueous solvents, a property known as "hydrophobicity". There are major energetic consequences of hydrophobicity for protein folding and in 1959 Kauzmann[5] proposed a solvent transfer model for quantifying them. The energetics of hydrophobicity are referred to here as the hydrophobic factor. Kauzmann argued that the free energy change for burying a non-polar side-chain in the interior of a protein can be found by experiments in which a model compound is partitioned between water and a non-aqueous solvent. Thus, the standard free energy change $\Delta G°$ for dissolving toluene in water gives an estimate for the free energy change that occurs on burying a phenylalanine side-chain inside a folded protein. The value of $\Delta G°$ for dissolving toluene in water at 25 °C is 22.8 kJ/mol:

$$\Delta G° = -RT\ln x \qquad (1)$$

where X is the solubility (mole fraction) of toluene in water. The assumptions made in analyzing solvent transfer experiments are discussed by Tanford.[6] Kauzmann gave estimates in the range −8 to −25 kJmol per side-chain for burying non-polar side-chains inside proteins. His early analysis[5] indicated that the hydrophobic factor would prove to be the most important energetic factor in protein folding.

In the following years, a large body of accurate experimental data was provided, first by Tanford's laboratory, whose work on this problem was summarized in a book,[6] and later by several workers, especially by Privalov and his laboratory. Basic questions were studied such as: which solutes will accurately model the transfer free energies of non-polar protein side-chains, which non-aqueous solvent best models the hydrophobic core of a protein, and how to take account of the van der Waals interactions made by side-chains buried inside proteins. The simplest experiments, for which the most accurate data are available, are ones in which the non-aqueous solvent is a neat liquid hydrocarbon and solution of the liquid hydrocarbon in water is studied. Protein chemists often require, however, transfer free energy data for the amino acid side-chains themselves, and data have been obtained for amino acids, blocked amino acids, and even pentapeptides.

A second major step towards quantifying the hydrophobic factor in protein folding was taken in the early 1970s when it was realized that the transfer free energy is likely to be proportional to the surface area of a nonpolar solute.[7–9] The water-accessible surface area (ASA) of a molecule is given by the Lee and Richards algorithm,[10] in which a spherical probe with a diameter appropriate for a water molecule is rolled over the solute's surface:

$$\Delta G_{\mathrm{hyd}} = k_{\mathrm{h}}(\mathrm{ASA}) \qquad (2)$$

After the proportionality coefficient k_h has been measured, the value of ΔG_{hyd} in a protein folding reaction can be computed from values of ASA for the native (N) and unfolded (U) forms of the protein:

$$\Delta G_{hyd} = k_h[(ASA)_N - (ASA)_U] \quad (3)$$

Different values of k_h have been proposed (see below), depending on questions such as which non-polar solvent should be used to model the protein interior, which solutes should be used to model the amino acid side-chains, and whether or not the values of ΔG_{hyd} should be corrected by the ratio of the solute/solvent molar volumes. The basic proposal that the transfer free energy is proportional to the buried apolar surface area is, however, increasingly viewed today as controversial (see below).

The packing–desolvation model

In Kauzmann's model the protein interior is treated as an organic liquid. There are reasons (see below) for doubting that this is a reliable assumption. An alternative model, referred to here as the "packing–desolvation" model, was proposed in 1988 by Privalov and Gill[11] and also by Ooi and Oobatake.[12] The focus originally[11] was on the apolar (ap) side-chains making close-packed van der Waals interactions in the hydrophobic core, but the polar peptide (p) groups also have an important energetic role both in the solvation and desolvation steps as well as in making or breaking peptide H-bonds:

$$\begin{array}{ccc} N & \rightarrow & U \ (vac) \\ \uparrow & & \downarrow \ (solv) \\ N & \rightarrow & U \ (aq) \end{array}$$

Scheme 1.

Scheme 1 relates the unfolding reaction in aqueous solution (N→U)(aq) to the (hypothetical) unfolding reaction in vacuum (N→U)(vac) by including steps for the desolvation of N, $-\Delta H_N$(solv), and the solvation of anhydrous U, ΔH_U(solv). Equating the two routes for unfolding N in aqueous solution gives (see equation (1) of Lazaridis *et al.*[4]):

$$\Delta H_{NU}(vac) = \Delta H_{NU}(aq) + \Delta H_{NU}(solv) \quad (4)$$

$$\Delta H_{NU}(solv) = \Delta H_N(solv) - \Delta H_U(solv) \quad (4a)$$

$$\Delta H_{solv} = \Delta H_p(solv) + \Delta H_{ap}(solv) \quad (4b)$$

The unfolding reaction in vacuum includes breaking van der Waals and H-bond interactions as well as an electrostatic term[4] arising from a difference in the dipole–dipole interactions in the peptide backbones of N and U.

The van der Waals or dispersion force interactions are well understood in model systems but it is not yet possible to test the calculations for proteins. Small alterations in distance between a pair of interacting atoms give large changes in predicted interaction energy and the computed energies are sensitive especially to the parameter R_{min}, the distance between atoms at the minimum in the potential energy curve. The average distances R_{ij} between interacting pairs of atoms (ij) of a given type (C–C, C–N, etc.) have been analyzed[13] in a set of 15 proteins. The interaction energy ΔE_{ij} between atoms i and j can be represented[4] by a Lennard–Jones potential as:

$$\Delta E_{ij} = \varepsilon[(R_{min}/R_{ij})^{12} - 2(R_{min}/R_{ij})^6] \quad (5)$$

To relate equation (5) to experimental enthalpy measurements, the assumption is made[4] that $\Delta E = \Delta H$. Most molecular mechanics force fields have similar values[4] for R_{min} and ε when expressing the interaction between a given pair of atoms. The 6th power term represents the attractive interaction between two atoms while the steeply rising 12th power term represents the repulsion that occurs at too close distances. Although the distances between atoms are widely distributed within a protein for any given type of atom pair, average distances are remarkably uniform[13] when results for the 15 proteins are compared and only distances larger than R_{min} are found. This finding is useful in developing an empirical calibration[13] that expresses how polar and non-polar groups contribute to measured enthalpies of protein unfolding.

Because proper evaluation of the van der Waals interaction energies in a protein folding reaction is still being discussed, the packing–desolvation model is not yet being used to quantify the energetics of non-polar side-chain burial, in contrast to Kauzmann's model. Hopefully this situation will change soon. As discussed below, there is considerable uncertainty about the best value of k_h to employ when using Kauzmann's model.

Molecular nature of hydrophobicity

A solute is hydrophobic if it is poorly soluble in water but readily soluble in non-aqueous solvents. Thus, mercury is not said to be hydrophobic. Because the solvation thermodynamics of a solute are defined[14] by experiments in which the solute is transferred from the vapor phase to liquid solution, the focus now shifts from liquid–liquid to gas–liquid transfer. A particular standard state,[14] namely 1 M solute in both the vapor and liquid phases, is used to eliminate the large and unwanted change in translational entropy between the vapor and liquid phases. Extensive thermodynamic experiments have been made (see below) on the transfer of hydrocarbon molecules from the gas phase to aqueous solution. Thus, ample data are available to discuss the energetics of desolvation of non-polar groups in the packing–desolvation model.

Solvation is modeled as a two-step process in Widom's solute insertion model.[15] In the first step,

thermal fluctuations in the arrangement of solvent molecules produce a cavity in the solvent, with a free energy change ΔG_c. In the second step, the solute enters the cavity, van der Waals interactions occur between solute and solvent, and rearrangement of the solvent occurs at the cavity surface; the solute–solvent interaction energy is denoted E_a.

Lee[16] has given quantitative values for the thermodynamic properties of the two steps in Widom's solute insertion model for some alkanes, including propane, isobutane and neopentane. These are flexible chain hydrocarbon molecules some of whose backbone conformations are restricted by branching: isobutane is singly branched and neopentane is doubly branched. Results for ΔG_c at 25 °C are shown in Figure 1 as a function of ASA for the solute. Table 1 gives values for E_a, obtained from Monte Carlo simulations[17] by Jorgensen and coworkers. The experimental transfer free energy $\Delta G_s°$ equals the sum of ΔG_c and E_a:

$$\Delta G_s° = \Delta G_c + E_a \quad (6)$$

The experimental $\Delta G_s°$ values were taken from vapor–liquid transfer data in the literature. Equation (6) is obtained by combining equations (15), (23) and (30) of Lee.[16] The E_a values obtained by simulation are believed to be highly reliable for these non-polar solutes and, because the $\Delta G_s°$ values are experimental, the ΔG_cvalues are termed "quasi-experimental".[16] E_a is assumed independent of temperature in deriving the ΔG_c values for water.

The results in Figure 1 and Table 1 show that the work of making a cavity is large in water compared to liquid alkanes. The difference is sufficiently large to explain why gaseous hydrocarbons have unfavorable solvation free energies in water although they have favorable values in liquid alkanes. When neopentane is dissolved in water, ΔG_c is 46.5 kJ/mol while a cavity of similar size in liquid neopentane costs only 27.6 kJ/mol. The interaction energy E_a for the van der Waals interactions (Table 1) includes the work of solvent reorganization when the solute enters the cavity. E_a has rather similar values when neopentane is dissolved either in water (E_a=−36.0 kJ/mol, assumed independent of temperature) or in liquid neopentane (E_a, which is temperature dependent, is −39.6 kJ/mol at 25 °C).

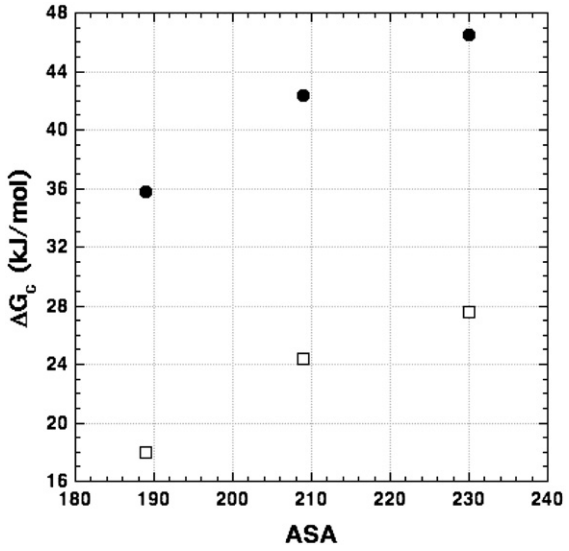

Figure 1. Comparison of the free energy of cavity formation in water (filled circles) and in liquid alkane (open squares) plotted against the water-accessible surface area (0.1 nm)^2 of the solute for propane, isobutane and neopentane. Data are at 25 °C from Lee.[16]

Table 1. Energetics at 25 °C of making a cavity in water compared to liquid alkane

Solute	ΔG_c	ΔH_c	$T\Delta S_c$	E_a
Water				
Propane	35.8	6.6	−29.2	−27.6
Isobutane	42.4	10.8	−31.6	−32.7
Neopentane	46.5	13.2	−33.3	−36.0
Liquid alkane				
Propane	23.7	40.1	16.4	−26.7
Isobutane	24.4	49.8	25.4	−35.1
Neopentane	27.6	12.4	15.2	−39

ΔG_c, ΔH_c and ΔS_c are the Gibbs energy, enthalpy, and entropy needed to make the cavity for the alkane solute, either in water or in liquid alkane; E_a is the van der Waals interaction energy between solute and solvent. All values are in kJ/mol at 25 °C; data are from Lee.[16] The liquid alkanes are maintained as liquids under their equilibrium vapor pressures.

Two effects contribute to the large work of making a cavity in water.[18] The first effect is the small size of the water molecule,[19] which makes it difficult to sculpt a cavity large enough for a neopentane solute because several water molecules must move in a concerted fashion to make room for one neopentane molecule. This first effect has been quantified[16] by using a hard sphere model of liquid solutions termed scaled particle theory (SPT). Comparison of SPT predictions with quasi-experimental values for thermodynamic properties of the two steps in solute insertion show surprisingly good agreement for most thermodynamic properties. This result implies that the small size of the water molecule makes a major contribution to the large work of making a cavity in water.

The second effect arises from the H-bonded structure of water. Because water has a large cohesive energy density (CED), it tends to squeeze out all solutes.[18,20] Polar solutes are soluble in water only because they make strong interactions with water molecules. In a simplistic approach one might ask which effect is dominant, the small size of the water molecule or the high CED of liquid water, by asking whether enthalpy or entropy dominates the ΔG_c for cavity formation, because CED is enthalpic while scaled particle theory, which favors the small size of the water molecule, is based on entropic considerations. At 25 °C, ΔG_c in water is dominated by entropy (Table 1) but the proportion of ΔG_c controlled by enthalpy rises with temperature[16] and, in fact, ΔG_c is controlled by ΔCp as is the solvation

free energy $\Delta G_s°$. Either the water H-bonds responsible for the large CED of water or the small size of the water molecule may be used to explain the unusual ΔCp: see below. For background reading on the relation between CED and hydrophobicity, see Graziano[21] and references therein.

Temperature dependence of hydrocarbon hydration

The energetics of hydrocarbon hydration (transfer of a gaseous hydrocarbon into water) depend on temperature in a highly specific and characteristic manner, and the temperature dependence is important in understanding the energetics of the desolvation–packing model for protein folding reactions. The values of $\Delta G°$, $\Delta H°$, $\Delta S°$ and ΔCp for neopentane hydration at 25 °C are shown in Table 2. The "signature" of hydrocarbon hydration is a large unfavorable (negative) entropy change at 25 °C coupled with a large positive value of ΔCp, the change in heat capacity for the transfer reaction. Note that the enthalpy change is negative and favors solvation.

As indicated by equation (7), ΔCp governs the temperature dependence of both $\Delta H°$ and $\Delta S°$:

$$\Delta C_p = d\Delta H°/dT = Td\Delta S°/dT \qquad (7)$$

For hydrocarbon hydration (gas→aqueous transfer), ΔCp decreases slowly with temperature from 5 °C to 130 °C.[22] The approximation that ΔCp is independent of temperature nevertheless occurs commonly in the literature, especially in discussing protein unfolding reactions for which this is a fairly good approximation for the protein between 0 °C and 60 °C.[23] As the temperature increases, both $\Delta H°$ and $T\Delta S°$ for hydrocarbon hydration become less negative while $\Delta S°$ goes to 0 above 100 °C. $\Delta G°$ (which is $\Delta H°–T\Delta S°$, and which is positive) increases slowly with temperature between 0 °C and 100 °C.[16,22]

For gas→aqueous transfer of non-polar solutes, the measured solvation enthalpy $\Delta H°_s$ is:[16]

$$\Delta H_s° = \Delta H_c + E_a \qquad (8a)$$

where ΔH_c is the enthalpy change for making a cavity in water for the hydrocarbon solute. When $\Delta H°_s=0$ (~95 °C for alkanes[22]), equation (8a) shows that $\Delta H_c=–E_a$, which is assumed to be temperature independent in water.[16] Thus, the solvation enthalpy may be expressed as the product of ΔCp and (95–T):

$$\Delta H_s°(T) = -\Delta Cp(95 - T) \qquad (8b)$$

When liquid→aqueous transfer data are analyzed, $\Delta H°=0$ at a much lower temperature ~25 °C.[24] This difference between the temperatures at which $\Delta H°=0$ (95 °C for gas→aqueous transfer of alkanes, 25 °C for liquid→aqueous transfer) represents an important energetic difference between the two transfer reactions. The difference arises because of the van der Waals interactions made by the alkane solute in liquid alkane.

A long-sought goal of work on the energetics of protein folding has been to construct a quantitative model that will predict $\Delta H°$, $\Delta S°$ and ΔCp of protein unfolding from energetic data on model systems, such as the solution of liquid hydrocarbons in water[24] or the transfer of gaseous hydrocarbons into water.[11,22,23,25] Reasonable success has been achieved in predicting ΔCp in protein unfolding from model compound data.[25,26] As mentioned above, there is a basic problem[13] in predicting $\Delta H°$ values for protein unfolding from model compound data because $\Delta H°$ represents a small difference between terms of opposite sign for disruption of H-bonds and van der Waals interactions on the one hand and hydration of newly exposed apolar and polar groups on the other.

The energetics of transferring gaseous neopentane into liquid neopentane[16] has very different thermodynamic properties from transfer into water, properties that are characteristic of "regular" solutions.[27] For transfer into liquid alkane, ΔCp and $T\Delta S°$ are small, both $\Delta H°$ and $T\Delta S°$ are nearly independent of temperature, and $\Delta G°$ is negative, i.e. favorable for transfer (Table 2). Neopentane makes favorable van der Waals interactions with both solvents, liquid alkane and water, but in water the solvation energetics contain a major contribution from ΔCp (see below). The values of $\Delta H°$ for transfer of a gaseous alkane into aqueous solution and into liquid alkane happen to be similar at 25 °C (Table 2), but they are quite different at 95 °C where the solvation enthalpy goes to zero for transfer into aqueous solution.[22]

The large positive value of ΔCp seen for hydrocarbon hydration (Table 2) is often explained by altered thermodynamics of the water molecules surrounding the dissolved hydrocarbon, primarily in the first solvation shell, as compared to bulk water. As the temperature increases, water molecules in the solvation shell gain increased freedom to behave like bulk water molecules, and this changing behavior can give rise to the large positive ΔCp value. Monte Carlo simulations[28] of the solvation thermodynamics for alkanes dissolved in water, using a potential function based on the random network model of water,[29] reproduce ΔCp semi-quantitatively. More surprisingly, scaled particle theory simulations of the thermodynamics of hydrocarbon solvation, which omit the water H-bonds, also reproduce ΔCp semi-quantitatively.[16]

Table 2. Comparison of the transfer energetics when neopentane is transferred at 25 °C from the gas phase either to liquid water or to liquid neopentane

Solvent	$\Delta H°$	$T\Delta S°$	$\Delta G°$	ΔCp
Water	−22.8	−33.3	10.5	363
Neopentane	−21.2	−9.2	−12.0	38

Contributions (kJmol) to the transfer free energy ($\Delta G°$) from enthalpy ($\Delta H°$) and entropy ($T\Delta S°$) when gaseous neopentane is transferred either to water or to liquid neopentane. The change in heat capacity ΔCp, which governs the temperature dependence of $\Delta H°$ and $\Delta S°$ (equation (6)), is given in J deg $^{-1}$ mol^{-1}. Values are referenced to the modified[16] Ben-Naim standard state[14] of 1 M in both the gas and liquid phases. Data are from Lee.[16]

Entropy convergence

An important property of hydrocarbon hydration is "entropy convergence". In 1979 Privalov[30] observed that the specific entropies (entropy per gram) of protein unfolding converge near 110 °C for several proteins when extrapolated linearly versus temperature. The solution entropies of liquid hydrocarbons in water show a related phenomenon: $\Delta S°$ goes to zero at a common temperature (T_s=113 °C)[24] when data taken at low temperatures are extrapolated linearly, which is equivalent to assuming ΔCp=constant (see below). If group additivity is valid, this observation indicates that the contribution of hydrophobic solvation to the entropy of protein unfolding may be removed by extrapolation to 113 °C. The protein unfolding data used for extrapolation should be taken at low temperatures (<60 °C) where ΔCp for protein unfolding is reasonably constant.[23] As mentioned above, modern studies of hydrocarbon hydration (gas to aqueous transfer) have been carried to high temperatures and show that ΔCp decreases with increasing temperature.[22] The temperature at which $\Delta S°$=0 is near 145 °C[22] when the decreasing nature of ΔCp is taken into account. Simulations of hydrocarbon hydration have been made that reproduce entropy convergence at T_s using either the information theory model[31] or scaled particle theory.[32]

Specific enthalpies of protein unfolding also converge[30] near 110 °C for several proteins when data taken at low temperatures are extrapolated linearly. When the solvation enthalpies of hydrocarbons are examined to investigate the underlying basis for this observation, the situation is found to be less well-defined. Aliphatic hydrocarbons show $\Delta H°$ going to zero at a common temperature T_h=95 °C but T_h is higher (about 125 °C) for aromatic hydrocarbons.[22] A modern survey of calorimetric data for protein unfolding[33] fails to confirm that the specific entropies and enthalpies converge near 110 °C, but the important point remains that data for hydrocarbons[24] show $\Delta S°$=0 at a common temperature T_s, which corresponds to 113 °C in a linear extrapolation.

Tests of the Kauzmann and packing–desolvation models

Attempts have been made to determine the value of ΔG_{hyd} directly from experiments on proteins in which "large to small" (e.g. Val→Ala) mutations are made and $\Delta\Delta G°$, the change in unfolding free energy produced by the mutation, is measured. The hope is that no change in protein structure or folding energetics occurs except for the decrease in buried non-polar ASA caused by truncating the side-chain, which gives an unfavorable entropy change. In a revealing set of experiments, the X-ray structures of the mutant proteins were measured as well as the $\Delta\Delta G°$ values for unfolding.[34] When the same Leu→Ala mutation was made at different sites in T4 lysozyme, cavities of different sizes were found. When the $\Delta\Delta G°$ values for unfolding were plotted against the size of each cavity and extrapolated back, the extrapolated $\Delta\Delta G°$ value at zero cavity size agreed[34] with the transfer free energy obtained using wet octanol as the non-aqueous solvent.[35]

When this type of mutational experiment was monitored, however, by scanning calorimetry, which gives not only the change in $\Delta G°$ but also the changes in $\Delta H°$ and $\Delta S°$ for unfolding, the results surprisingly showed favorable entropy changes.[36] This finding indicates that the dominant entropy change caused by the mutation does not come from the decrease in buried non-polar ASA but instead comes from increased conformational entropy. Consequently, the $\Delta\Delta G°$ values for unfolding cannot be assigned wholly to changes in buried non-polar surface area. It remains a challenging problem to test Kauzmann's model with mutational experiments on proteins.

Concern about neglecting van der Waals interactions when using Kauzmann's model arose especially from the observation[37] that side-chains buried in the protein interior are close-packed like atoms in an organic crystal. Close packing represents about a 10% decrease in volume compared to the organic liquid, and the dependence of attractive van der Waals interactions on distance between interacting atoms indicates that substantially stronger interaction energies must be present in the protein interior than in an organic liquid. When it became possible to redesign the hydrophobic cores of proteins and to predict protein stability for different amino acid sequences, a comparison was made[38] of the relative energetic roles of van der Waals interactions and burial of non-polar surface area. The results indicate that close-packed van der Waals interactions do indeed increase protein stability.[38] If instead the protein interior should resemble a non-aqueous solvent, as predicted by Kauzmann's model, then the side-chains should not be close-packed and van der Waals interactions should have little effect on protein stability. The van der Waals interactions would then be nearly the same for non-polar side-chains in the folded and unfolded states, since they have similar values for hydrocarbon solutes dissolved either in water or liquid alkane (Table 1).

The suggestion has been made[39] that van der Waals interactions can be neglected and burial of non-polar surface is sufficient to account for the folding energetics of non-polar side-chains even when they are close-packed, because melting of hydrocarbon crystals produces only small changes in free energy, since the entropic and enthalpic changes nearly cancel. As discussed above, computer design studies indicate that van der Waals interactions do contribute markedly to the thermal stability of proteins.[38]

Choice of k_h values when using Kauzmann's model

Kauzmann's model leaves unanswered the basic question of which non-aqueous solvent best

approximates the hydrophobic core of a protein. Protein chemists often quote ΔG_{hyd} values based on a solvent transfer system in which the non-aqueous solvent is "wet" (water-saturated) *n*-octanol.[35,40] However, wet octanol contains 2.3 M H_2O[41] and in 1988 Radzicka and Wolfenden[41] found transfer free energies between water and wet cyclohexane (which is nearly anhydrous) that are more than twice as large as those between water and wet octanol.

For this and other reasons, it is not possible to say with confidence that any specified values for ΔG_{hyd} and k_h in equations (2) and (3) are the "right" values. Some widely cited values are as follows. In 1974 Chothia[8] gave the value of k_h=9.2 kJ mol^{-1} nm^{-2}, based on published ΔG_{hyd} values from Tanford's laboratory for the solubilities of non-polar amino acids in ethanol and dioxane. Chothia drew separate lines[8] for the non-polar amino acids and ones that contain H-bonding side-chain groups. In 1983 Fauchère and Pliska[35] provided revised ΔG_{hyd} values for the amino acids based on partitioning of blocked amino acids between water and wet *n*-octanol. Earlier ΔG_{hyd} values, measured by using amino acids without blocked end groups, gave erroneous ΔG_{hyd} values because the charged end groups interfere with using the group additivity principle to analyze the data. In 1996 Wimley *et al*.[40] repeated the partitioning experiments between water and wet octanol using pentapeptides rather than blocked amino acids. They found significant changes in the ΔG_{hyd} values when pentapeptides are studied, and they attributed this result to "occlusion" (reducing the solvent exposure) of the side-chain under study by neighboring side-chains. They give a value for k_h of 10.0 kJ mol^{-1} nm^{-2}.

In 1991 Sharp and co-workers[42] suggested that the ΔG_{hyd} values from solvent transfer experiments on hydrocarbon solutes should be corrected for the ratio of solute to solvent molar volumes; a similar correction is known in polymer chemistry as the "Flory–Huggins" correction. Using a series of linear alkanes and data for their solubilities in water, they gave a value for k_h of 13.0 kJ mol^{-1} nm^{-2} for liquid–liquid transfer without the volume correction, when the partition coefficients on the mole fraction scale are used (Figure 2), as compared to k_h=19.7 kJ mol^{-1} nm^{-2} with the volume correction.[42] Whether or not the volume correction should be made has been debated extensively. This is the main subject of a lengthy review by Chan and Dill[43] who recommend against making the correction and who recommend (for other reasons) using a k_h value of 13.0 kJ mol^{-1} nm^{-2}.

Proportionality between hydrophobic free energy and ASA

The basic assumption that ΔG_{hyd} is proportional to ASA rests on empirical calibrations such as the one shown in Figure 2, which plots ΔG_{hyd} versus ASA for liquid-aqueous transfer of linear alkanes (the number of carbon atoms, n=1–10).[42] There are

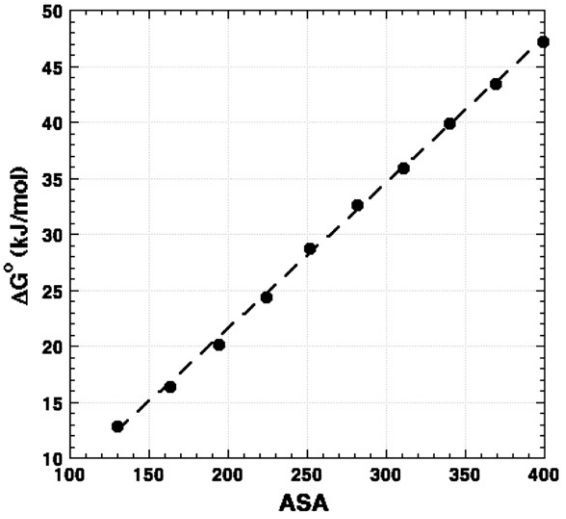

Figure 2. Linear dependence of transfer free energy on ASA (0.1 nm)^2 for linear alkanes (from one to ten carbon atoms) for transfer from the neat liquid hydrocarbon to water at 25 °C (see the text). Data are from Sharp *et al*.[42] $\Delta G°$ is $-RT$ ln X, where X is the solubility (mol fraction) of the hydrocarbon in water The slope is 13.0 kJ/nm^2 and the line does not go through (0,0).

good reasons, however, to believe that proportionality between ΔG_{hyd} and ASA is only a rough approximation. Gas to aqueous transfer data for saturated hydrocarbons show big deviations from a single correlation line when branching, linear and cyclic hydrocarbons are included in the same plot.[44] The plot shown in Figure 2 is in effect a linear plot of ΔG_{hyd} *versus* the number of carbon atoms (see similar plots byTanford[6] and Ben-Naim and Marcus[14]), when the ASA values are calculated for extended alkane chains. But long, flexible alkane molecules tend to curl up into more spherical shapes in aqueous solution, and the plot in Figure 2 might be curved if the ASA values could be derived for the appropriate average conformation actually present in aqueous solution. The theoretical basis for a possible proportionality between ΔG_{hyd} and ASA has been analyzed recently[45] and the authors conclude that the assumed proportionality is unsatisfactory in accurate work.

Pairwise interaction between hydrocarbon molecules in water

The mutual interaction between a pair of hydrocarbon molecules dissolved in water has been of particular interest to theoretical chemists. In 1977 Pratt and Chandler studied the problem by an analytic approach,[46] with results that stimulated much further work on simulation of the pairwise interaction both by Monte Carlo and molecular dynamics methods. When the pairwise interaction energy is plotted *versus* distance between two hydrocarbon molecules, two energy minima are seen, one for the contact interaction and one for the solvent-separated interaction.

However, the interaction energy for pairwise interaction between hydrocarbon molecules in water is known experimentally to be extremely weak, and consequently its role in protein folding is obscure. Dimer formation by a pair of benzene molecules in water is barely detectable: the dissociation constant has been measured at 25 °C to be 0.035 mol fraction.[47]

When Monte Carlo simulations are made of a system containing several hydrocarbon molecules in a box of 216 water molecules, contact clusters of hydrocarbon molecules are formed and cluster formation is cooperative.[48] The apparent equilibrium constant for adding a hydrocarbon molecule to a contact cluster is independent of cluster size only above a critical size and smaller clusters show weaker association behavior. In the size-independent range the free energy change for adding a solute molecule to a contact cluster gives[48] a value of k_h (equation (2)) equal to 10.0 kJ mol^{-1} nm^{-2}, in agreement with values from solvent transfer experiments. Solvent-separated clusters are found before the contact clusters appear, suggesting that desolvation of the solvent-separated clusters may be a measurable kinetic step in the protein folding process.[49]

Further reading on the hydrophobic factor

(1) The hydrophobic factor is important in analyzing various problems besides protein folding, notably problems involving membranes. Blokzijl and Engberts[50] have provided a monumental and clearly written review of the hydrophobic factor as it affects a whole range of problems. (2) Dill and co-workers[51] have used a simplified model of water, the "Mercedes-Benz" model, to interpret the hydrophobic factor. This water model, which emphasizes the hydrogen bonding properties of water, allows simulation studies to be made that otherwise would require excessive computer time. (3) Chandler[52] and co-workers have analyzed the transition that occurs for cavity formation in water when the cavity size changes from large to small. For large cavities, ΔG_c is given by the simple expression γA, where γ is the surface tension of water and A is the surface area of the cavity, whereas for small cavities the temperature dependence of ΔG_c differs from that of γ.[52] (4) Hummer co-workers[53] have analyzed cavity formation in water with simulation studies and have shown that an information-theory model of cavity sizes provides a remarkably simple and useful representation. (5) Further background reading on the hydrophobic factor in protein folding is given in an earlier review[54] by me. (6) Stites and co-workers[55] have gathered mutational data on staphylococcal nuclease aimed at testing whether van der Waals interactions in the protein core stabilize folding. They find that thermal stability of the protein correlates well with the number of van der Waals contact interactions. (7) Israelachvili and co-workers[56] summarize years of work on the interaction between two macroscopic flat plates in water as a function of the distance between plates. A chief goal of this work is to obtain a first principles perspective on the hydrophobic factor.

Peptide H-bonds and peptide solvation

The twin subjects of peptide H-bonds and peptide solvation are closely connected because H-bonds between water and the peptide NH and CO groups must be broken before peptide H-bonds can be formed. Peptide desolvation during folding is the largest single energetic factor in folding when desolvation is considered separately from formation of peptide H-bonds. However, peptide desolvation caused by folding is always an exchange reaction accompanied by H-bonding of the buried peptide NH and CO groups. This is shown both by modern reanalysis[57] of protein structures that were initially reported to contain buried but non-H-bonded polar groups and by an experimental study[58] of Ala→Ser and Val→Thr mutations in T4 lysozyme. The newly introduced polar groups[58] were always found to be H-bonded either to water molecules or to protein polar groups.

Interpreting measurements of $\Delta H°$ or $\Delta G°$ for making a peptide H-bond in aqueous solution necessarily includes desolvation of the NHCO group as well as making the peptide H-bond. Nevertheless, theoretical treatments of peptide H-bonds and peptide solvation follow different lines. Quantum chemistry applied to molecules in the gas phase provides most current knowledge of peptide H-bond energies while current work on peptide solvation is largely based on electrostatic calculations that treat water as a continuum solvent. The reason is that the H-bonded structure of water is too complex for treatment by quantum mechanics when water is represented as a random network of H-bonds. The first serious treatments of peptide solvation by quantum mechanics[59–61] have been appearing recently, however, and further developments are likely to come soon.

Important progress has been made on using titration calorimetry to measure $\Delta H°$ values of peptide H-bonds in peptide helices. These results leave little doubt that peptide H-bonds are a major energetic factor in protein folding, and they connect the energetics of peptide H-bonds to the propensities of the different amino acids to form α-helices. In the following, experimental work on peptide H-bonds is discussed first and then current theoretical efforts to understand peptide H-bonds and peptide solvation.

The energetics of peptide helices

In 1989 the surprising finding was made[62] that the peptide H-bond clearly stabilizes peptide helices in water, because alanine forms a helix without the aid of any helix-stabilizing interactions such as salt bridges. The term "peptide H-bond" is used here to include the associated van der Waals and long-range electrostatic interactions discussed by Lazaridis *et*

al.[4] in their dissection of the helix-forming reaction. Because the side-chain of alanine is just a methyl group, the special ability of alanine helix to form a stable helix strongly implies that the peptide H-bond itself is helix-stabilizing in water. For solubility in water, helical alanine peptides require the admixture of a polar amino acid (Lys^+, Glu^- and Gln have been used), but a long sequence of 13 uninterrupted alanine residues is soluble when two basic amino acids are added at each end, and the alanine sequence forms a helix.[63] Alanine is the only one of the 20 naturally occurring amino acids known to form a stable peptide helix[64] in water unassisted by helix-stabilizing interactions. Absolute helix propensities, not just values relative to a reference amino acid, can be measured in an alanine-based peptide host and alanine has a helix propensity of 1.70.[64]

Titration calorimetry shows that alanine helix formation is enthalpy-driven. The value for ΔH is $-3.8(\pm 0.4)$ kJ/mol per residue over the temperature range 5 °C to 45 °C[65–67] and ΔCp is too small to measure. Helix enthalpy data have been obtained recently for several amino acids by Makhatadze and co-workers.[67] β-Branched amino acids, such as valine and isoleucine, have somewhat smaller ΔH values clustered around −2.5 kJ/mol, and amino acids with H-bonding side-chains, such as serine and threonine, also have $\Delta H = -2.5$ kJ/mol, while leucine and phenylalanine have values around −3.8 kJ/mol, close to that of alanine.[67] Thus, peptide helix formation is enthalpy-driven in water for all amino acids studied while ΔH varies with the side-chain. Because ΔCp is very small, these results cannot be explained by burial of non-polar surface area in the helix, which would require a substantial positive value of ΔCp for helix unfolding instead of the observed negligibly small value. The side-chains contribute to the different helix propensities of the amino acids[67] *via* the loss in side-chain conformational entropy[68] that occurs on helix formation.

Thus, peptide helix results show that solvent-exposed peptide H-bonds are stabilizing for protein folding. If buried peptide H-bonds contribute equally strongly as solvent-exposed ones, then the energetic role of peptide H-bonds in protein folding can be estimated as follows. A typical 100-residue protein, which has ~70 peptide H-bonds, should gain: $\Delta H = \sim(70)(-3.2) = -225$ kJ/mol. A ΔH value of −3.2 kJ/mol per H-bond is used as an average for the different amino acids (see above). For comparison, burial of non-polar surface area, estimated by Kauzmann's model, should contribute to the folding of a 100-residue protein as follows: $\Delta G = -(100)(0.50)(10.5) = -525$ kJ/mol. On average, each residue buries 0.50 nm^2 of non-polar surface as there is a linear relation between the number of residues in a protein and the amount of buried non-polar area.[33] For the choice of k_h=10.5 kJ mol^{-1} nm^{-2}, see above. Thus, peptide H-bonds, with their associated van der Waals interactions, contribute substantially to the energetics of protein folding if the ΔH value for solvent-exposed peptide H-bonds is applicable.

Paradox of enthalpy-driven helix formation

There is a basic paradox[69] concerning the fact that peptide helix formation is enthalpy-driven. Three[60] H-bonds to water (one to the peptide NH and two to the CO group) are broken when a peptide H-bond is made, according to a quantum chemistry study[60] of the solvation of *N*-methylacetamide (NMA). On the other hand, only one peptide H-bond per residue is made in an α-helix or β-sheet. This paradox can be stated quantitatively by using amides to model the solvation enthalpy of the free NHCO group: the value given by Makhatadze and Privalov[3] is −59.6 kJ/mol at 25 °C. The energy of the peptide H-bond (gas phase, various model compounds) has been estimated[70] by quantum chemistry as $-25(\pm 4)$ kJ/mol, and the enthalpy should be somewhat smaller.[60] Thus, desolvation of the free NHCO group should cost 59.6 kJ/mol[3], based on calorimetric data for amides, while less than −25 kJ/mol is gained by making a peptide H-bond, and the enthalpy deficit is more than 34 kJ/mol. Nevertheless, alanine peptide helix formation is enthalpy-driven with $\Delta H° = -3.8$ kJ/mol per residue (see above) and there must be either an error or an omission in this statement of the paradox.

A plausible resolution of the paradox is given by electrostatic calculations of peptide solvation,[71–74] as well as by quantum chemical calculations,[61] which indicate that amides are unsatisfactory models for the solvation enthalpy of the NHCO group in peptides. These calculations are described below.

In the older literature this paradox is avoided by the "H-bond inventory",[75] which is based simply on counting the H-bonds made and consumed as peptide H-bonds are made and by including water–water H-bonds in the inventory. The H-bond inventory fails[69] to predict the solvation enthalpies of amides whose experimental values[76] are known accurately, and the H-bond inventory is rarely used today.

Electrostatic calculations of peptide solvation

For a long time, solvation of the peptide group was considered to be a fixed quantity[3] that is independent of peptide backbone conformation. In this view, the solvation enthalpy of the peptide group can be taken directly from experimental data for simple amides by assuming group additivity. It is necessary to subtract the contribution from any non-polar groups[3,77] and also to make a standard state correction[77] for transfer of the amide from the gas phase to liquid solution. Electrostatic calculations of peptide solvation[71–74] indicate, however, that the solvation enthalpy of the NHCO group is far from being a constant and the principle of group additivity is invalid when applied to the polar NH and CO groups. Because dipole–dipole interactions in the peptide backbone determine the solvation energy, its value changes with peptide backbone conformation and neighboring side-chains also affect peptide solvation by controlling the access of water to the

peptide group. The values for the solvation enthalpy of the NHCO group in peptides found by electrostatic calculations[71–74] are almost twofold smaller than the value used by Makhatadze and Privalov.[3,77] This difference between values used for the solvation enthalpy of the NHCO group is a key factor in explaining why Lazaridis *et al*.[4] find different values for the enthalpies of protein unfolding than Makhatadze and Privalov[3] (see below).

Electrostatic calculations of peptide solvation have been made[71–74] by using a standard electrostatic algorithm (DelPhi[78]) to solve the Poisson–Boltzmann equation, using continuum electrostatics. The basic premise is that the large partial charges on the peptide NH and CO groups are responsible for the solvation of these polar groups. Fixed, atom-centered partial charges are used. The geometry of the molecule under study must be known, but otherwise there are no adjustable parameters.

A different electrostatic algorithm, based on a Langevin dipoles model of water,[79] has also been used to calculate solvation free energies of peptides.[73] When the results were compared with DelPhi results, agreement within ±4 kJ/mol was found,[73] which is typical of comparative results found using different theoretical methods for calculating solvation energies[80] (see below). Recently comparison has been made[80] of a wide range of theoretical methods for predicting the known solvation free energies of model compounds. The various methods include different levels of quantum theory as well as force field calculations that employ either explicit or implicit solvent. Using either explicit solvent or a high level of quantum theory fails to improve the agreement[80] with experiment significantly. Apparently some basic problem remains to be solved[80] such as finding a suitable water model for calculating solvation energies. The results of several methods give agreement within ±4 kJ/mol.

Comparison of experimental results for solvation free energies and enthalpies of amides show that the free energy of solvating the NHCO group is almost entirely enthalpic (±4 kJ/mol)[71] and consequently the calculated value of electrostatic solvation free energy (ESF) values) gives approximately the solvation enthalpy of the NHCO group in amides and peptides. Although it is surprising that the solvation entropy of the NHCO group in amides is small compared to the solvation free energy, the reason is simply that the solvation enthalpy is large. The same situation[74] is found for the relative free energy, enthalpy and entropy of hydration of simple ions.[81] ESF values are used to explain the paradox of enthalpy-driven helix formation in the following section.

The ESF value of the NHCO group in peptides depends on the peptide backbone conformation because the relative alignment of the NHCO dipoles in neighboring residues varies strongly with backbone conformation. For example, neighboring NHCO dipoles are antiparallel in the extended-β conformation but they are parallel in the α-helical conformation.[82] Because ESF depends strongly on backbone conformation, the ESF of the NHCO group should play a major role in determining α-helix and β-structure propensities of the amino acids. The relation between ESF and β-structure propensity was studied[72] for variants of the zinc finger protein used to determine β-structure propensities.[83] The ESF values of peptide groups adjacent to the substitution site in the protein, which lies within a β-hairpin, change significantly with the substitution. There is a linear correlation[72] between ESF and β-structure propensity, which was determined[83] from the change in free energy of unfolding.[83] However, the correlation is inverse,[72] meaning that the amino acid substitution affects the ESF values of peptide groups in the unfolded protein more than those in the folded protein.

ESF values were also used to investigate[84] the "blocking effect"[85] observed when peptide hydrogen exchange rates are compared for amino acids with different non-polar side-chains. Both the blocking effect and the ESF of the NHCO group should depend on the group's access to solvent and a linear correlation is indeed found[84] between the blocking effect and the ESF value of the NHCO group.

Unfolding enthalpies in vacuum: role of the solvation enthalpy of the peptide group

In 1995 a major effort was made to compare theory with experiment for the energetics of protein folding. Makhatadze and Privalov[3] (MP) summarized the extensive studies from their laboratory of the unfolding enthalpies of four proteins. As a first step in the interpretation of results, they removed the effects of solvation by using model compound data[77] for both polar and apolar groups. Lazaridis, Archontis and Karplus[4] (LAK) then simulated the unfolding enthalpies in vacuum of the four proteins by using the molecular force field CHARMM. Even at this first stage of analysis, there was a major discrepancy between the quasi-experimental and simulated values: the simulated values were approximately twofold smaller than the quasi-experimental ones. LAK[4] then investigated a simpler system, the alanine peptide helix, and the twofold discrepancy persisted. In computing the unfolding enthalpy in vacuum by the MP[3] procedure, desolvation of the polar peptide groups makes a much larger contribution than any other factor (see the breakdown in Table 3) taken from LAK.[4]

A possible explanation for the ~2-fold discrepancy between the quasi-experimental and simulated values (F. Avbelj and R.L.B., unpublished results) became apparent upon comparing ESF values for the NHCO group in peptides versus monoamides.[71,74] As explained above, the ESF of the NHCO group in an amide is approximately equal to its solvation enthalpy and the same is assumed to be true in peptides: then the solvation enthalpy can be estimated from the calculated ESF of the peptide group. The MP[3,77] value for the solvation enthalpy of the NHCO group is −59.6 kJ/mol (25 °C) while the ESF of the NHCO group in an alanine peptide (polypro-

Table 3. Protein unfolding enthalpies in vacuum found by the Makhatadze & Privalov[3] procedure

Protein	ΔH_{NU}(aq)	ΔH_{ap}(solv)	ΔH_{p}(solv)	ΔH_{NU}MP
cyt *c*	88	1029	6381	7498
RNase A	297	916	7757	8966
Lys	243	1092	7402	8736
Mb	6	1573	9397	10,976
Protein	ΔH_{NU}LAK	MP/LAK	ΔH^{*}_{p}(solv)	MP*/LAK
cyt *c*	3946	1.90	4064	1.31
RNase A	4468	2.01	4941	1.38
Lys	4669	1.87	4714	1.30
Mb	6242	1.76	5986	1.21

Upper half: breakdown into component parts of the unfolding enthalpies in vacuum of the four proteins studied by Makhatadze & Privalov.[3] Columns: (1) the measured unfolding enthalpy in aqueous solution; (2) the apolar solvation enthalpy; (3) the polar (peptide) solvation enthalpy; and (4) the overall unfolding enthalpy in vacuum, the sum of columns 1, 2 and 3. Lower half: columns: (1) the simulated solvation enthalpy in vacuum of LAK[4]; (2) the ratio of the MP/LAK values for the unfolding enthalpy in vacuum; (3) the polar solvation enthalpy reduced by 1.57, as suggested by ESF calculations (see the text); and (4) the ratio of the MP/LAK values for the unfolding enthalpy in vacuum when the polar solvation enthalpy is reduced by 1.57. The four proteins are: cyt *c*, horseheart cytochrome *c*; RNase A, bovine pancreatic ribonuclease A; lys, hen eggwhite lysozyme; and Mb, sperm whale myoglobin. All entries are in kJ/mol at 25 degrees C.

line II conformation[86]) is only −38 kJ/mol;[84] the ratio of the two values is 1.57. Table 3 shows the effect of reducing by 1.57 the estimated contribution from the peptide groups to the vacuum unfolding enthalpy: this reduces from ~2-fold to ~1.3-fold the average ratio of the two values for ΔH_{NU}(vac).

The alanine peptide helix, discussed by LAK[4], gives a more direct view of the role of the solvation enthalpy of the NHCO group because the amount of apolar surface buried in the alanine helix is small, so that the apolar contribution to the solvation enthalpy is almost negligible, and also because the two polar terms (solvation of the helix and desolvation of the unfolded peptide) can be taken into account individually. For the method of obtaining the unfolding enthalpy in vacuum, see Scheme 1 and equations (4)–(4b) above. LAK[4] applied the MP[3] procedure for solvating N and desolvating U and obtained $\Delta H_{NU,p}$(solv)=43 kJ/mol, $\Delta H_{NU,ap}$(solv)=2 kJ/mol, giving ΔH_{NU}(solv)=43+2=45 kJ/mol (see their Table XIII and equation (4b), above). When the modern calorimetric value for ΔH_{NU}(aq) of 3.8kJ/mol[65–67] is used, ΔH_{NU}(vac)=45+4=49 kJ/mol per residue (see equation (4), above). This quasi-experimental value of 49 kJ/mol is nearly twofold larger than the simulated value[4] of 27 kJ/mol per residue for the unfolding enthalpy of the alanine helix in vacuum.

Better agreement is obtained between the quasi-experimental and simulated values for ΔH_{NU}(vac) if ESF values for the alanine helix and unfolded peptide (together with ΔH_{p}(solv)=ESF) are used to solvate N and desolvate U. The ESF of the NHCO group in the interior of an alanine helix is −10 kJ/mol[71] and its ESF is −38 kJ/mol in the unfolded peptide[84] (polyproline II conformation[86]). Then $\Delta H_{NU,p}$(solv)=−10+38=28 kJ/mol, ΔH_{NU}(solv)=28+2=30 kJ/mol, ΔH_{NU}(aq)=3.8 kJ/mol and ΔH_{NU}(vac)=30+4=34 kJ/mol per residue.

These ESF values also explain fairly well the paradox of enthalpy-driven helix formation, discussed above. The problem now is to take an assumed value for the peptide H-bond enthalpy in the gas phase (−25 kJ/mol;[70] see paradox statement above) and use the ESF-derived values above for the other quantities in Scheme 1 to find ΔH_{NU}(aq)=ΔH_{NU}(vac)−ΔH_{NU}(solv). This gives ΔH_{NU}(solv)=30 kJ/mol, ΔH_{UN}(vac)=−25 kJ/mol and ΔH_{NU}(aq)=−25+30=5 kJ/mol. This predicted value of ΔH_{NU}(aq) is 9 kJ/mol larger than the measured value of −4 kJ/mol, while the predicted value given in the paradox is >34 kJ/mol. A likely explanation for the 9 kJ/mol discrepancy is that the alanine helix is stabilized not only by the peptide H-bond but also by associated van der Waals interactions[4] and possibly also by long-range electrostatic interactions in the peptide backbone.[4]

Failure of the principle of group additivity for the polar NH and CO groups

It has been known for a long time that the principle of group additivity may not be valid for the polar NH and CO groups (see, for example, the 1988 paper by Roseman[87]). As discussed above, ESF values for the NHCO group in monoamides and peptides[71] revealed that the predicted solvation free energy of the NHCO group is more negative in monoamides than in peptides and the predicted failure of group additivity has been examined.[74] Experimental values for the solvation enthalpy of the NHCO group in monoamides are close (±4 kJ/mol) to the amide ESF values,[71] and so the same prediction could be made for the solvation enthalpy of the NHCO group, namely that it is substantially more negative in monoamides than in peptides. In principle this prediction can be tested experimentally. The solvation enthalpy of a crystalline amide is measured by combining the results of two experiments, the heat of solution and the heat of sublimation, and it has been difficult to obtain the heat of sublimation for peptides.

In 2006 Della Gatta[88] and co-workers succeeded in getting calorimetric data for the solvation enthalpies of four dipeptides. They find that the average solvation enthalpy of the two NHCO groups in dipeptides, divided by the average for the NHCO group in seven *N*-alkylamides, is 0.76. This is a direct demonstration of the failure of group additivity. Comparison between the measured solvation enthalpies of the dipeptides and their predicted ESF values is forthcoming (F. Avbelj and R.L.B., unpublished results). Further evidence that group additivity fails for the polar NH and CO groups is provided by a calorimetric study of substituted ureas.[89] The measured solvation enthalpy of the HN-CO-NH group in the substituted ureas is −71.3 kJ/mol[89]

while the value estimated by summing the solvation enthalpies of the isolated NH and CO groups is −96.3 kJ/mol,[89] which is 1.35-fold larger. This result further confirms the failure of group additivity.

Peptide H-bonds

A simple electrostatic model of the peptide H-bond emerged from the pioneering force field studies of Lifson and co-workers.[90] In this model the peptide H-bond is represented by the two dipoles on the peptide NH and CO groups. Because the dipole–dipole interaction is most favorable when the two dipoles are collinear, the H-bond is predicted to be linear with an 180° angle between the C=O group and facing N atom. This simple electrostatic representation is well suited to rapid computation with modern force fields, and it is widely used today.[91] However, peptide H-bonds deviate in some respects from the predictions of the electrostatic model[91,92] and there is much current interest in developing more accurate models.[93] A key question is whether peptide H-bonds are formed cooperatively in protein folding: for example, whether the H-bond energy in an α-helix increases in going from either helix end towards the center. Most of the literature on this subject contains simulations of anhydrous helices and β-hairpins in the gas phase.

Quantum mechanical calculations for the gas phase show that the energies of different backbone conformations change drastically when the peptide backbone is solvated. The polyproline II (P_{II}) conformation is the most stable conformation of the alanine dipeptide (acetyl-alanine-*N*-methylamide) when the Ala dipeptide is H-bonded to four water molecules, but P_{II} is not even among the stable conformations in the absence of water.[59] Moreover, there is a large decrease in the energy difference between any two stable conformations of the alanine dipeptide in the gas phase when it is H-bonded to four water molecules.[59] The energy difference between unsolvated dipeptide conformations found by quantum mechanics is well correlated with the local backbone energy difference found by classical electrostatics,[94] indicating that electrostatics accounts for a substantial part of the overall conformational energy difference.

Crystal structures of H-bonded chains of formamide and urea molecules with varying chain lengths provide a key system for understanding amide H-bonds in crystals. The structural results from crystals have been analyzed by Dannenberg,[92] using density functional theory (DFT), a quantum mechanical calculation. A basic structural result is that the H-bond length varies with the position of the monomer in these chains, whose chain lengths vary from 2 to 15 monomers.[92] DFT analysis shows that the H-bonds are formed cooperatively: when crystal structures of varying chain lengths are compared, the average energy per monomer increases and the average H-bond length decreases with increasing number of monomers in the amide chain.[92] The H-bond enthalpy varies from −19 kJ/mol in the formamide dimer to −54 kJ/mol in the central H-bond of a chain of 15 formamide units.[92]

The energetics of alanine peptide helices in the gas phase have been analyzed by a DFT procedure similar to the one used to fit crystal structures of formamide chains, and cooperative properties of helix formation have been predicted.[92,95] Approximately one-half of the predicted cooperativity of helix formation in the gas phase results from redistribution of the electron densities of atoms forming the H-bonds and half comes from long-range electrostatic interactions, when cooperativity is defined by the increase in H-bond energy with position in the helix.[95]

It is important to obtain a better H-bond model for use with molecular mechanics force fields in order to represent the electron redistribution that occurs when peptide H-bonds are formed. Polarizable force fields are being developed for this purpose.[93] Electron redistribution in peptide H-bonds has been observed in a high-resolution X-ray structure of cholesterol oxidase.[96] Polarization of the –C=O groups of α-helices is observed while the –C=O groups of β-sheets show electron density shared between the C and O atoms. This difference between the electron redistribution in α-helices and β-sheets probably explains[97] the opposite signs observed for the secondary structure chemical shifts assigned to α-helices and β-sheets.

Interestingly, gas phase calculations for peptide helices and β-sheet structures have been interpreted as showing strain energy in alanine α-helices in the gas phase, as much as 29 kJ/mol per residue for long helices, while H-bonds in β-sheets appear to be less strained.[92] According to PDB structural data, the average H-bond angle (the angle between the C=O group and facing N atom) is 153(±2)° in protein helices and has the same mean value in antiparallel β-sheets,[98] while the angle is somewhat larger in parallel β-sheets (157°–162°, depending on the number of connecting β-strands[98]). These values are significantly different from the 180° angle predicted by the simple electrostatic model. The close agreement between H-bond angles for α-helices and β-sheets suggests that the difference between 153° and 180° does not arise from strain. In contrast to results for α-helices, the average H-bond angle for 3_{10} helices is strikingly different from 153° and drops from 122° to 106° as the number of helical turns increases from one to five turns,[98] and this effect indicates the H-bonds are strained in 3_{10} helices. In PDB structures of proteins, the measured H-bond length in α-helices does change somewhat with position in the helix. The H-bond length decreases slightly near the helix centers compared to the ends,[98] but the effect is smaller than found for predicted peptide helices in the gas phase.[92,98]

A recent DFT analysis by Baker and co-workers of the energetics of forming the amyloid structure[99] provides important insight into the role of H-bonds in forming the amyloid spine. The amyloid structures formed by various six and seven-residue peptides have been determined by X-ray crystal-

lography[100] and one high-resolution structure has been subjected to DFT analysis.[99] An important structural feature is that the cross-β amyloid spine is anhydrous.[100] Each amyloid layer contains two peptide molecules and the polar seven-residue peptide GNNQQNY gives an amyloid layer structure in which there are 22 H-bonds to a neighboring layer: 12 peptide backbone H-bonds and ten side-chain H-bonds per peptide dimer.[100] The H-bonds in conjunction with van der Waals interactions are sufficiently strong to draw adjacent amyloid layers close together without intervening water molecules.

The DFT energy value per peptide dimer does not give information on the individual contributions of H-bonds, long-range electrostatic interactions, and van der Waals interactions, but estimates of these factors have been made by classical procedures (D. Baker, personal communication). The DFT value for the total interaction energy per amyloid layer (containing a peptide dimer) in an infinite crystal has the impressive value −841 kJ/mol.[99] The total energy in the infinite crystal estimated by classical electrostatics is −502 kJ/mol, giving −23 kJ/mol per H-bond, while the estimated van der Waals energy per peptide dimer is −251 kJ/mol (D. Baker, personal communication). Thus the sum of the van der Waals energy plus H-bond energy estimated by classical electrostatics is somewhat less negative (−753 kJ/mol) than the total energy found by quantum chemistry (−841 kJ/mol). The authors compare the average total energy per H-bond (−38 kJ/mol) with the energy of a water H-bond in ice (−28 kJ/mol) and with the average energy per H-bond when only two amyloid layers, containing two pairs of peptides, interact (−33 kJ/mol). The latter comparison (−38 kJ/mol for the infinite crystal *versus* −33 kJ/mol for just two layers) illustrates the energetic cooperativity of forming this amyloid structure. Data for the equilibrium enthalpy and free energy differences between the amyloid structure and the unfolded peptide are not yet available.

Although calorimetric data[65–67] show clearly that solvent-exposed peptide H-bonds stabilize folding, it has been controversial[101,102] whether or not buried peptide H-bonds also contribute favorably to protein stability. Resolution of the controversy may come from considering the close-packing[103,104] produced by van der Waals interactions, since close packing could strengthen peptide H-bonds by allowing the H-bonding NH and CO groups to draw closer together. The fact that strong H-bonds are observed in an amyloid structure[99] that has neither buried non-polar side-chains nor buried water within the amyloid spine[100] certainly indicates that buried peptide H-bonds, can stabilize folding. It appears that just the cooperativity of forming a network of H-bonds is sufficient to stabilize this amyloid structure.

Backbone conformational entropy

The backbone conformational entropy is by far the largest single factor in the energetics of folding that opposes the folding process. It is the only known energetic factor that has a large unfavorable value in all conditions. Because the net ΔG value favoring folding of globular proteins is small, between −20 kJ/mol and −60 kJ/mol,[105] in comparison to the −525 kJ/mol contribution estimated (see above) for the burial of non-polar surface area in a 100-residue protein, the free energy change from backbone entropy should be nearly equal, but opposite in sign, to the sum of all the other factors that contribute to the free energy of folding.

The change in backbone conformational entropy upon folding can be estimated from W, the number of available conformations per residue, in the unfolded peptide if these conformations are chosen at random. It is assumed the folded peptide has just one available backbone conformation:

$$\Delta S_{\mathrm{res}} = R \ln W \qquad (10)$$

In a pioneering study of peptide folding energetics in 1955, John Schellman[106] gave estimates for an upper bound (W=36, ΔS_{res}=−30 J deg^{-1} mol^{-1}) and a lower bound (ΔS_{res}=−12 J deg^{-1} mol^{-1}), while omitting the effects of side-chains. The lower bound was based on the conformational entropy of the C–C bond in alkanes. Today it is known that each of the two backbone rotatable bonds (ϕ,ψ) has three principal conformations (*trans*, *gauche*$^-$, *gauche*$^+$), and so a first estimate of W=3×3=9. Then ΔS_{res}= −18.3 J deg^{-1} mol^{-1} and ΔG_{res}=5.46 kJ/mol at 25 °C. However, the backbone conformations do not occur at random in unfolded peptides, and certain amino acids have surprisingly strong backbone conformational preferences. For example, alanine is reported to occur chiefly in the polyproline II conformation.[86] These conformational preferences will cause $-\Delta S_{\mathrm{res}}$ to be smaller than above. Note that the estimate of ΔG_{res} is roughly equal but opposite in sign to the value given above (−5.25 kJ/mol) for the average free energy change per residue caused by burial of non-polar surface area in a folded protein.

The effect of the side-chain on backbone conformational entropy has been analyzed in a careful simulation study made by Amzel co-workers.[107] They find ΔS_{res}= −17 J deg^{-1} mol^{-1} for Ala, −27 J deg^{-1} mol^{-1} for Gly, and −14 J deg^{-1} mol^{-1} for all other amino acids except Val and Ile, which have values of −9 J deg^{-1} mol^{-1}. Thus, for Gly there is a considerable increase in the number of available backbone conformations caused by the absence of a β–carbon atom. Other effects that may contribute to ΔS_{res}, such as excluded volume, were found to be small effects.[107]

The total conformational entropy change upon protein unfolding has been estimated[24] by taking the experimental value for the unfolding entropy change and subtracting the contribution from buried non-polar surface area. This subtraction is possible because model compound data show that the entropy of solvation of different non-polar compounds goes to zero at a common high temperature (T_s), estimated[24] to be 113 °C, when a linear extrapolation is used (see above). For hen lysozyme, this method of estimating the total conformational entropy change

gave[24] $\Delta S_{res}=-2280/129=-17.7$ J deg^{-1} mol^{-1}. Comparing this value for the total change in conformational entropy with estimates of the backbone entropy change (see above) indicates that other factors contributing to the total entropy change, such as side-chain entropy, should be small.

A possible important contribution from the solvation entropy of the polar peptide groups has not been considered here. As discussed above, it should be small if the experimental values of ΔG and ΔH for solvation of the NHCO group in monoamides[71] provide a reliable guide; they indicate that $\Delta G \sim \Delta H$ and $T\Delta S \ll \Delta G$. Amzel and co-workers[107] agree that the "polar entropy" is small while Makhatadze and Privalov[3] consider it important.

Both the enthalpy and free energy of forming an alanine peptide helix are known experimentally for alanine-based peptides and ΔS_{res} can be calculated from $\Delta G=\Delta H-T\Delta S$: then ΔS_{res} has the surprisingly low value[86] of -9 J deg^{-1} mol^{-1}. This value includes, however, an entropy change of opposite sign from burial of non-polar surface area in the helix, which contributes a free energy change at 25 °C of -1.9 kJ/mol[4]. Correction for this effect gives $\Delta S_{res}=-15$ J deg^{-1} mol^{-1} for alanine.

The overall conclusion concerning how backbone entropy changes on folding is that generally agreed-on values for different amino acids are not yet available, but probably ΔS_{res} is close to -17 J deg^{-1} mol^{-1} for alanine while most other amino acids, except Gly, should have lower values, as discussed by Amzel[107] and co-workers.

Specific pairwise interactions, helix and β-structure propensities, disulfide bonds, and cooperativity of folding

Pairs of side-chains make various types of specific interactions in proteins such as salt bridges, H-bonds and cation–pi interactions. An individual salt bridge can contribute as much as -17 kJ/mol to the stability of a folded protein.[108] This is highly significant compared to the net stability of a folded protein,[105] but it is rare for a small protein to have more than a few such strong pairwise interactions and they are not reviewed here, although a few comments are made. The strong pairwise interaction mentioned above is a H-bonded salt bridge formed between oppositely charged side-chains of an Asp and a His residue in T4 lysozyme.[108] Mutating one residue to Asn and then using ^{1}H-NMR to measure the pK change of the other residue was used to give the energetics of breaking the salt bridge.

Because Coulombic interactions are long-range and other charged residues besides the salt bridge pair can contribute, the double mutant cycle[109] has been particularly helpful in analyzing salt bridges. A recent comparison of a salt bridge in ubiquitin with the reverse salt bridge illustrates the usefulness of this approach; results from other salt bridge studies are also given.[110] One reason for the wide interest in how Coulombic interactions affect folding stability is that, being long-range, charge–charge interactions may be established at early stages in the folding process.[111] Another reason is that multiple salt bridges are frequently found in the thermophilic forms of certain proteins, especially oligomeric forms.[112]

Side-chain to side-chain H-bonds between pairs of polar residues occur fairly often in proteins (there are about 11 such H-bonds per 100 residues[113]), and they occur especially within α-helices. The H-bond strength for a Gln·Asp (i, i+4) H-bond has been measured in a peptide helix system as $\Delta G°=-4.2$ kJ/mol when the Asp residue is charged and -1.6 kJ/mol when it is uncharged.[114]

A large number of specific pairwise interactions are included as variable parameters in the algorithm AGADIR of Muñoz and Serrano,[115] together with parameters for helix propensities and related factors. Statistical comparison between observed and predicted helix contents of a large number of partly helical peptides was used to give the values of the pairwise interaction parameters and other variables.[115] Several later refinements of the original procedure have been published.

Helix and β-structure propensities have been studied extensively using various peptide hosts for substitution experiments and the helix propensities are in good agreement with values measured in proteins where other energetic factors, such as burial of non-polar surface, are difficult to separate out. Much is known about helix and β-structure propensities, but only a few comments are made here. The factors believed responsible for secondary structure propensities involve both major factors (e.g. peptide H-bonds) and auxiliary factors (e.g. side-chain entropy) in the energetics of folding. A basic property of helix propensities is that helix formation is enthalpy driven and the favorable enthalpy is provided chiefly by the peptide H-bond (see above). A second main factor determining helix propensities[67] is the loss of side-chain conformational entropy[68] on helix formation. Another basic property is that helix propensities are context-dependent, meaning that their values depend on the choice of host peptide. For example, the differences between helix propensities of non-polar amino acids are twice as large in alanine-based peptides as those measured in a mixed-sequence peptide from RNase T1.[116] Nevertheless, there is good linearity when the helix propensity differences from the two systems are plotted against each other.[116] A straightforward explanation for this type of context dependence is that helix and β-structure propensities depend on the ESF values of peptide groups adjacent to the substituted side-chain[71,72] because the ESF change for Ala → Val is substantially larger in an alanine-based peptide than in a mixed-sequence peptide like the RNase T1 peptide.

Helix propensities measured in the RNase T1 peptide have the same values, relative to a reference amino acid, as those measured in the RNase T1 protein itself[116] when a solvent-exposed position is used as the substitution site This finding from Scholtz[116] and co-workers provides reassurance about using helix propensities measured in peptide helices to help analyze protein folding energetics.

Context dependence affects both β-structure and helix propensities as well as β-structure propensities. β-Sheets are more complex structures than α-helices and side-chain interactions between β-strands obviously affect the results of substitution experiments. A synthetic peptide template has been developed recently[117] that has advantageous properties for measuring the large number of possible side-chain to side-chain interactions that are involved. In general, differences in β-structure propensity among the amino acids are smaller than differences in helix propensity, and the order of helix propensities tends to be opposite to the order of β-structure propensities.

The peptide NH groups of the first four residues of an isolated α-helix are not included in the NH groups forming helical peptide H-bonds and likewise the last four CO groups of a helix are not included. Special "helix cap" structures are found commonly at both ends of protein helices that compensate (or sometimes over-compensate) for these H-bond deficiencies.[118] Gly and Pro destabilize both α-helices and β-structures, although they can be helix-stabilizing in specific helix cap structures.[118]

Covalent cross-links that connect two segments of the polypeptide backbone stabilize folding A commonly occurring cross-link is the disulfide bond (-S–S-) formed by mutual oxidation of two cysteine residues. Proteins that contain disulfide bonds typically unfold when they are reduced. The conformational entropy of the polypeptide backbone in an unfolded protein decreases when a disulfide bond is formed. This is a large energetic effect, often sufficient to explain the stabilizing effect of a disulfide bond, although enthalpic effects may also occur. There is a linkage relation between stabilization of folding and disulfide bond stability that has been studied by Creighton:[119] the folded structure of a protein increases the stability of an internal –S–S- bond exactly as the disulfide bond increases the stability of folding. A clear example of how this principle operates to promote folding is given by the unfolded polypeptide chain of bovine pancreatic trypsin inhibitor (BPTI).[120] A reverse turn that contains a five-residue hydrophobic cluster promotes formation of the 14–38 –S–S-bond and *vice versa*.[120] This example illustrates how cooperative interactions play a basic role in the folding process:[119,120] each successive interaction (step *i*) stabilizes the preceding ones (*i*-1, *i*-2...) and the earlier interactions make possible the formation of step *i*.

Coulombic interactions

Globular proteins contain ionizing residues that are partly or fully exposed to solvent, both acidic (Asp, Glu, Tyr) and basic (Lys, Arg, His) residues. The charges on the ionized forms of these residues interact according to Coulomb's law and typically the charge–charge interaction energy becomes large and positive (unfavorable) both at low pH, where only the basic residues are ionized and there is a net positive charge, and at high pH, where ionized acidic residues dominate and there is a net negative charge. A strong influence of charge–charge interactions on protein stability is seen when T_m, the temperature midpoint of thermal unfolding, is plotted against pH: in typical cases T_m drops sharply as the pH decreases from 7 to 2. Anion binding contributes to the net charge and is likely to contribute substantially to the plot of T_m *versus* pH, which is determined in good part simply by the variation of net charge with pH. The participation of anion binding becomes evident[121] when these plots are measured at a series of salt concentrations and compared for a weakly binding anion such as Cl^- *versus* a strongly binding anion such as SO_4^{2-}.

There is much interest in using methods of protein engineering to increase protein stability, and Coulombic interactions provide a useful tool.[122,123] A simple electrostatic model (a variant of the Tanford–Kirkwood model) has been used to predict roughly the changes in stability caused by mutations that involve charged residues.[122] When the relation between charge–charge interactions and stability is examined in detail, using methods such as directed evolution to increase protein stability, it is apparent that the problem is not yet fully solved.[123]

Protein stability depends on pH *via* differences between the residue p*K* values in the native and unfolded forms of a protein. pH titration of the ionizing residues can be used to determine the change in protein stability between two pH values provided that both the native protein and the unfolded protein (usually in a strong denaturant such as GdmCl) can be titrated.[124] The relation was tested for RNase A by comparing the change in net proton charge of the protein between two pH values with the change in protein stability measured by denaturant-induced unfolding, and good agreement was found.[124] This approach was used to quantify[124] the folding free energy change at pH 3.0, where there is a large positive net charge, when NaCl is added to reduce the Coulombic interaction energy. Until recently it has been standard practice to assume that the p*K* values of the unfolded protein can be estimated from p*K* data for model peptides. This procedure fails, however, when there is significant structure in the "unfolded protein" because structure affects the p*K* values.[125]

Acknowledgements

I thank Franc Avbelj, Joe Dannenberg, B.-K. Lee, Themis Lazaridis, George Makhatadze, and George Rose for discussion and Joe Dannenberg and David Eisenberg for sending me their manuscripts before publication.

References

1. Anfinsen, C. B. (1973). Principles that govern the folding of protein chains. *Science*, **181**, 223–230.

2. Dill, K. A. (1990). Dominant forces in protein folding. *Biochemistry*, **29**, 7133–7155.
3. Makhatadze, G. I. & Privalov, P. L. (1995). Energetics of protein structure. *Adv. Protein Chem.* **47**, 307–425.
4. Lazaridis, T., Archontis, G. & Karplus, M. (1995). Enthalpic contribution to protein stability: insights from atom-based calculations and statistical mechanics. *Adv. Protein Chem.* **47**, 231–306.
5. Kauzmann, W. (1959). Factors in interpretation of protein denaturation. *Adv. Protein Chem.* **14**, 1–63.
6. Tanford, C. (1980). In *The Hydrophobic Effect.* 2nd edit. John Wiley and Sons, New York.
7. Hermann, R. B. (1972). Theory of hydrophobic bonding. II. The correlation of hydrocarbon solubility in water with cavity surface area. *J. Phys. Chem.* **76**, 2754–2759.
8. Chothia, C. (1974). Hydrophobic bonding and accessible surface area in proteins. *Nature*, **248**, 339.
9. Reynolds, J. A., Gilbert, D. B. & Tanford, C. (1974). Empirical correlation between hydrophobic free energy and aqueous cavity surface area. *Proc. Natl Acad. Sci. USA*, **71**, 2925–2927.
10. Lee, B. & Richards, F. M. (1971). The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* **55**, 379–400.
11. Privalov, P. L. & Gill, S. J. (1988). Stability of protein structure and hydrophobic interaction. *Adv. Protein Chem.* **39**, 191–234.
12. Ooi, T. & Oobatake, M. (1988). Effects of hydrated water on protein unfolding. *J. Biochem. (Tokyo)*, **103**, 114–120.
13. Hilser, V. J., Gómez, J. & Freire, E. (1996). The enthalpy change in protein folding and binding: refinement of parameters for structure-based calculations. *Proteins: Struct. Funct. Genet.* **26**, 123–133.
14. Ben-Naim, A. & Marcus, Y. (1984). Solvation thermodynamics of nonionic solutes. *J. Chem. Phys.* **81**, 2016–2027.
15. Widom, B. (1982). Potential distribution theory and the statistical mechanics of fluids. *J. Phys. Chem.* **86**, 869–872.
16. Lee, B. (1991). Solvent reorganization contribution to the transfer thermodynamics of small non-polar molecules. *Biopolymers.* **31**, 993–1008.
17. Jorgensen, W. L., Gao, J. & Ravimohan, C. (1985). Monte Carlo simulations of alkanes in water. Hydration numbers and the hydrophobic effect. *J. Phys. Chem.* **89**, 3470–3473.
18. Lazaridis, T. (2001). Solvent size vs cohesive energy as the origin of hydrophobicity. *Acc. Chem. Res.* **34**, 931–937.
19. Lee, B. (1985). The physical origin of the low solubilities of non-polar solutes in water. *Biopolymers.* **24**, 813–823.
20. Tanford, C. (1979). Interfacial free energy and the hydrophobic effect. *Proc. Natl Acad. Sci. USA*, **76**, 4175–4176.
21. Graziano, G. (2004). Relationship between cohesive energy density and hydrophobicity. *J. Chem. Phys.* **121**, 1878–1882.
22. Makhatadze, G. I. & Privalov, P. L. (1994). Energetics of interactions of aromatic hydrocarbons with water. *Biophys. Chem.* **50**, 285–291.
23. Privalov, P. L. & Makhatadze, G. I. (1992). Contribution of hydration and noncovalent interactions to the heat capacity effect on protein unfolding. *J. Mol. Biol.* **224**, 715–723.
24. Baldwin, R. L. (1986). Temperature dependence of the hydrophobic interaction in protein folding. *Proc. Natl Acad. Sci. USA*, **83**, 8069–8072.
25. Makhatadze, G. I. & Privalov, P. L. (1993). Contribution of hydration to protein folding thermodynamics. I. The enthalpy of hydration. *J. Mol. Biol.* **232**, 639–659.
26. Gómez, J., Hilser, V. J., Xie, D. & Freire, E. (1995). The heat capacity of proteins. *Proteins: Struct. Funct. Genet.* **22**, 404–412.
27. Hildebrand, J. H. & Scott, R. L. (1962). *Regular Solution.* Prentice-Hall, Englewood Cliffs, NJ.
28. Sharp, K. A. & Madan, B. (1997). Hydrophobic effect, water structure, and heat capacity changes. *J. Phys. Chem. ser. B*, **101**, 4343–4348.
29. Henn, A. R. & Kauzmann, W. (1989). Equation of state of a random network, continuum model of liquid water. *J. Phys. Chem.* **93**, 3770–3783.
30. Privalov, P. L. (1979). Stability of proteins: small globular proteins. *Adv. Protein Chem.* **33**, 167–241.
31. Garde, S., Hummer, G., Garcia, A., Paulaitis, M. E. & Pratt, L. R. (1996). Origin of entropy convergence in hydrophobic hydration and protein folding. *Phys. Rev. Letters*, **77**, 4966–4968.
32. Graziano, G. & Lee, B.-K. (2003). Entropy convergence in hydrophobic hydration: a scaled particle theory analysis. *Biophys. Chem.* **105**, 241–250.
33. Robertson, A. D. & Murphy, K. P. (1997). Protein structure and the energetics of protein stability. *Chem. Rev.* **97**, 1251–1267.
34. Eriksson, A. E., Baase, W. A., Zhang, X.-J., Heinz, D. W., Blaber, M., Baldwin, E. P. & Matthews, B. W. (1992). Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic effect. *Science*, **255**, 178–183.
35. Fauchère, J.-L. & Pliska, V. (1983). Hydrophobic parameters π of amino acid side-chains from partitioning of N-acetyl-amino-acid-amides. *Eur. J. Med. Chem.* **18**, 369–375.
36. Loladze, V. V., Ermolenko, D. M. & Makhatadze, G. I. (2002). Thermodynamic consequences of burial of polar and nonpolar amino acid residues in the protein interior. *J. Mol. Biol.* **320**, 343–357.
37. Richards, F. M. (1977). Areas, volumes, packing and protein structure. *Annu. Rev. Biophys. Bioeng.* **6**, 151–176.
38. Dahiyat, B. I. & Mayo, S. L. (1997). Probing the role of packing specificity in protein design. *Proc. Natl Acad. Sci. USA*, **94**, 10172–10177.
39. Nicholls, A., Sharp, K. A. & Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins: Struct. Funct. Genet.* **11**, 281–296.
40. Wimley, W. C., Creamer, T. P. & White, S. H. (1996). Solvation energies of amino acid side-chains and backbone in a family of host-guest peptides. *Biochemistry*, **35**, 5109–5124.
41. Radzicka, A. & Wolfenden, R. (1988). Comparing the polarities of the amino acids: side-chain distribution coefficients between the vapor phase, cyclohexane, 1- octanol, and neutral aqueous solution. *Biochemistry*, **27**, 1644–1670.
42. Sharp, K. A., Nicholls, A., Friedman, R. & Honig, B. (1991). Extracting hydrophobic free energies from experimental data: relationship to protein folding and theoretical models. *Biochemistry*, **30**, 9686–9697.
43. Chan, H. S. & Dill, K. A. (1997). Solvation: how to obtain microscopic energies from partitioning and solvation experiments. *Annu. Rev. Biophys. Biomol. Struct.* **26**, 425–459.
44. Simonson, T. & Brünger, A. T. (1994). Solvation free energies estimated from macroscopic continuum

theory: an accuracy assessment. *J. Phys. Chem.* **98**, 4683–4694.
45. Wagoner, J. A. & Baker, N. (2006). Assessing implicit models for nonpolar mean solvation forces: the importance of dispersion and volume terms. *Proc. Natl Acad. Sci. USA*, **103**, 8331–8336.
46. Pratt, L. R. & Chandler, D. (1977). Theory of the hydrophobic effect. *J. Chem. Phys.* **67**, 3683–3704.
47. Tucker, E. E., Lane, E. H. & Christian, S. D. (1981). Vapor pressure studies of hydrophobic interactions. Formation of benzene-benzene and cyclohexane-cyclohexanol dimers in dilute aqueous solution. *J. Solut. Chem.* **10**, 1–20.
48. Raschke, T. M., Tsai, J. & Levitt, M. (2001). Quantification of the hydrophobic interaction by simulation of the aggregation of small hydrophobic solutes in water. *Proc. Natl Acad. Sci. USA*, **98**, 5965–5969.
49. Rank, J. A. & Baker, D. A. (1997). Desolvation barrier to cluster formation may be a rate-limiting step in protein folding. *Protein Sci.* **6**, 347–354.
50. Blokzijl, W. & Engberts, J. B. F. N. (1993). Hydrophobic effects. Opinons and facts. *Angew. Chem. Int. Ed. Engl.* **32**, 1545–1579.
51. Southall, N. T., Dill, K. A. & Haymet, A. D. J. (2002). A view of the hydrophobic effect. *J. Phys. Chem. B*, **106**, 521–533.
52. Chandler, D. (2005). Interfaces and the driving force of hydrophobic assembly. *Nature*, **437**, 640–647.
53. Hummer, G., Garde, S., García, A., Pohorille, A. & Pratt, L. R. (1996). An information theory model of hydrophobic interactions. *Proc. Natl Acad. Sci. USA*, **93**, 8951–8955.
54. Baldwin, R. L. (2005). Weak interactions in protein folding: hydrophobic free energy, van der Waals interactions, peptide hydrogen bonds, and peptide solvation. In *Protein Folding Handbook* (Buchner, J. & Kiefhaber, T., eds), Part I. *Wiley-V.C.H.* Weinheim.
55. Chen, J., Lu, Z., Sakon, J. & Stites, W. E. (2000). Increasing the thermostability of staphylococcal nuclease: implications for the origin of thermostability. *J. Mol. Biol.* **303**, 125–130.
56. Meyer, E. E., Rosenberg, K. J. & Israelachvili, J. (2006). Recent progress in understanding hydrophobic interactions. *Proc. Natl Acad. Sci. USA*, **103**, 15739–15746.
57. Fleming, P. J. & Rose, G. D. (2005). Do all backbone polar groups in proteins form hydrogen bonds? *Protein Sci.* **14**, 1911–1917.
58. Blaber, M., Lindstrom, J. D., Gassner, N., Xu, J., Heinz, D. W. & Matthews, B. W. (1993). Energetic cost and structural consequences of burying a hydroxyl group within the core of a protein determined from Ala→Ser and Val→Thr substitutions in T4 lysozyme. *Biochemistry*, **32**, 11363–11373.
59. Han, W.-G., Jalkanen, K. J., Elstner, M. & Suhai, S. (1998). Theoretical study of aqueous *N*-acetyl-L-alanine-*N*′-methylamide: structures and Raman, VCD and ROA spectra. *J. Phys. Chem. ser. B*, **102**, 2587–2602.
60. Dannenberg, J. J. (2006). Enthalpies of hydration of N-methylacetamide by one, two and three waters and the effect upon the C=O stretching frequency. An *ab initio* study. *J. Phys. Chem. ser. A*, **110**, 5798–5802.
61. Salvador, P. Asensio, A. & Dannenberg, J. J. (2007). The effect of aqueous solvation upon a-helix formation for polyalanines. *J. Phys. Chem. ser. A*. In the press.
62. Marqusee, S., Robbins, V. H. & Baldwin, R. L. (1989). Unusually stable helix formation in short alanine-based peptides. *Proc. Natl Acad. Sci. USA*, **86**, 5286–5290.
63. Spek, E. J., Olson, C. A., Shi, Z. & Kallenbach, N. R. (1999). Alanine is an intrinsic α-helix stabilizing amino acid. *J. Am. Chem. Soc.* **121**, 5571–5572.
64. Rohl, C. A., Chakrabartty, A. & Baldwin, R. L. (1996). Helix propagation and N-cap propensities of the amino acids measured in alanine-based peptides in 40 volume percent trifluoroethanol. *Protein Sci.* **5**, 2623–2637.
65. Lopez, M. M., Chin, D.-H., Baldwin, R. L. & Makhatadze, G. I. (2002). The enthalpy of the alanine peptide helix measured by isothermal titration calorimetry using metal-binding to induce helix formation. *Proc. Natl Acad. Sci. USA*, **99**, 1298–1302.
66. Goch, G., Maciejczyk, M., Oleszczuk, M., Stachowiak, D., Malicka, J. & Bierzynski, A. (2003). Experimental investigation of initial steps of helix propagation in model peptides. *Biochemistry*, **42**, 6840–6847.
67. Richardson, J. M., Lopez, M. M. & Makhatadze, G. I. (2005). Enthalpy of helix-coil transition: missing link in rationalizing the thermodynamics of helix-forming propensities of the amino acid residues. *Proc. Natl Acad. Sci. USA*, **102**, 1413–1418.57.
68. Creamer, T. P. & Rose, G. D. (1994). α-Helix-forming propensities in peptides and proteins. *Proteins: Struct. Funct. Genet.* **19**, 85–97.
69. Baldwin, R. L. (2003). In search of the energetic role of peptide H-bonds. *J. Biol. Chem.* **278**, 17581–17588.
70. Mitchell, J. B. O. & Price, S. L. (1991). On the relative strengths of amide···amide and amide···water hydrogen bonds. *Chem. Phys. Letters*, **180**, 517–523.
71. Avbelj, F., Luo, P. & Baldwin, R. (2000). Energetics of the interaction between water and the helical peptide group and its role in determining helix propensities. *Proc. Natl Acad. Sci. USA*, **97**, 10786–10791.
72. Avbelj, F. & Baldwin, R. L. (2002). Role of backbone solvation in determining thermodynamic β–propensities of the amino acids. *Proc. Natl Acad. Sci. USA*, **99**, 1309–1313.
73. Avbelj, F. (2000). Amino acid conformational preferences and solvation of polar backbone atoms in peptides and proteins. *J. Mol. Biol.* **300**, 1335–1359.
74. Avbelj, F. & Baldwin, R. L. (2006). Limited validity of group additivity for the folding energetics of the peptide group. *Proteins: Struct. Funct. Bioinformat.* **63**, 283–289.
75. Fersht, A. R. (1987). The hydrogen bond in molecular recognition. *Trends Biochem. Sci.* **12**, 301–304.
76. Della Gatta, G., Barone, G. & Elia, V. (1986). Enthalpies of solvation for some N-alkylamides in water and in carbon tetrachloride at 25 °C. *J. Solut. Chem.* **15**, 157–167.
77. Makhatadze, G. I. & Privalov, P. L. (1993). Contribution of hydration to protein folding thermodynamics. I. The enthalpy of hydration. *J. Mol. Biol.* **232**, 639–659.
78. Sitkoff, D., Sharp, K. A. & Honig, B. (1994). Accurate calculation of hydration free energies using macroscopic solvent models. *J. Phys. Chem.* **98**, 1978–1988.
79. Florian, J. & Warshel, A. (1997). Langevin dipoles model for ab initio calculations of chemical processes in solution: parameterization and application to hydration free energies of neutral and ionic solutes and conformational analysis in aqueous solution. *J. Phys. Chem. ser. B*, **101**, 5583–5595.
80. Mobley, D. L., Dumont, E., Chodera, J. D. & Dill, K. A. (2007). Comparison of charge models for fixed-charge force fields: small-molecule hydration free

energies in explicit solvent. *J. Phys. Chem. ser. B*, **111**, 2242–2252.
81. Robinson, R. A. & Stokes, R. H. (1955). *Electrolyte Solutions.* Butterworths, London.
82. Avbelj, F. & Moult, J. (1995). Role of electrostatic screening in determining protein main chain conformational preferences. *Biochemistry*, **34**, 755–764.
83. Kim, C. A. & Berg, J. (1993). Thermodynamic β-sheet propensities measured using a zinc finger host peptide. *Nature*, **362**, 267–270.
84. Avbelj, F. & Baldwin, R. L. (2004). Origin of the neighboring residue effect on peptide backbone conformation. *Proc. Natl Acad. Sci. USA*, **101**, 10967–10972.
85. Bai, Y. & Englander, S. W. (1994). Hydrogen bond strength and β-sheet propensities: the role of a side-chain blocking effect. *Proteins: Struct.Funct. Genet.* **18**, 262–266.
86. Shi, Z., Olson, C. A., Rose, G. D, Baldwin, R. L. & Kallenbach, N. R. (2002). Polyproline II structure in a sequence of seven alanine residues. *Proc. Natl Acad. Sci. USA*, **99**, 9190–9195.
87. Roseman, M. A. (1988). Hydrophilicity of polar amino acid side-chains is markedly reduced by flanking peptide bonds. *J. Mol. Biol.* **200**, 513–522.
88. Della Gatta, G., Usacheva, T., Badea, E., Palecz, B. & Ichim, D. (2006). Thermodynamics of solvation of some small peptides in water at T=298.15 K. *J. Chem. Thermodynam.* **38**, 1054–1061.
89. Della Gatta, G., Badea, E., Jozwiak, M. & Del Vecchio, P. (2007). Thermodynamics of solvation of urea and some monosubstituted *N*-alkylureas in water at 298.15 K. *J. Chem. Eng. Data*, **52**, 419–425.
90. Lifson, S., Hagler, A. T. & Dauber, P. (1979). Consistent force-field studies of intermolecular forces in hydrogen-bonded crystals. I. Carboxylic acids, amides and the C=O···H-hydrogen bonds. *J. Am. Chem. Soc.* **101**, 5111–5121.
91. Morozov, A. V. & Kortemme, T. (2006). Potential functions for hydrogen bonds in protein structure prediction and design. *Adv. Protein Chem.* **72**, 1–38.
92. Dannenberg, J. (2006). The importance of cooperative interactions and a solid-state paradigm to proteins: what peptide chemists can learn from molecular crystals. *Adv. Protein Chem.* **72**, 227–273.
93. Friesner, R. A. (2006). Modeling polarization in proteins and protein-ligand complexes: methods and preliminary results. *Adv. Protein Chem.* **72**, 79–104.
94. Avbelj, F., Grdadolnik, S. G., Grdadolnik, J. & Baldwin, R. L. (2006). Intrinsic backbone preferences are fully present in blocked amino acids. *Proc. Natl Acad. Sci. USA*, **103**, 1272–1277.
95. Morozov, A. V., Tsemekhman, K. & Baker, D. (2006). Electron density redistribution accounts for half the cooperativity of α-helix formation. *J. Phys. Chem. ser. B*, **110**, 4503–4505.
96. Lario, P. I. & Vrielink, A. (2003). Atomic resolution density maps reveal secondary structure dependent differences in electronic distribution. *J. Am. Chem. Soc.* **125**, 12787–12794.
97. Avbelj, F., Kocjan, D. & Baldwin, R. L. (2004). Protein chemical shifts arising from α-helices and β-sheets depend on solvent exposure. *Proc. Natl Acad. Sci. USA* **101**, 17394–17397.
98. Koch, O., Bocola, M. & Klebe, G. (2005). Cooperative effects in hydrogen-bonding of protein secondary structure elements: a systematic analysis of crystal data using Secbase. *Proteins: Struct. Funct. Genet.* **61**, 310–317.
99. Tsemekhman, K., Goldschmidt, L., Eisenberg, D. & Baker, D. (2007). Cooperative hydrogen-bonding in amyloid formation. *Protein Sci.* **16**, 761–764.
100. Nelson, R., Sawaya, M. R., Balbirnie, M., Madsen, A. Ø., Riekel, C., Grothe, R. & Eisenberg, D. (2005). Structure of the cross-β spine of amyloid-like fibrils. *Nature*, **435**, 773–778.
101. Honig, B. & Yang, A.-S. (1995). Free energy balance in protein folding. *Adv. Protein Chem.* **46**, 27–58.
102. Pace, C. N., Treviño, S., Prabhakaran, E. & Scholtz, J. M. (2004). Protein structure, stability, and solubility in water and other solvents. *Proc. Trans. Roy. Soc. ser. B*, **359**, 1225–1235.
103. Honig, B. (1999). Protein folding: from the Levinthal paradox to structure prediction. *J. Mol. Biol.* **293**, 283–293.
104. Schell, D., Tsai, J., Scholtz, J. M. & Pace, C. N. (2006). Hydrogen bonding increases packing density in the protein interior. *Proteins: Struct. Funct. Bioinformat.* **63**, 278–282.
105. Pace, C. N. (1990). Conformational stability of globular proteins. *Trends Biochem. Sci.* **15**, 14–17.
106. Schellman, J. A. (1955). The stability of hydrogen-bonded peptide structures in aqueous solution. *C.R. Trav. Lab. Carlsberg, Sér. Chim.* **29**, 230–259.
107. D'Aquino, J. A., Gómez, J., Hilser, V. J., Lee, K. H., Amzel, L. M. & Freire, E. (1996). The magnitude of the backbone conformational entropy change in protein folding. *Proteins: Struct. Funct. Genet.* **25**, 143–156.
108. Anderson, D. E., Becktel, W. J. & Dahlquist, F. W. (1990). pH-induced denaturation of proteins: a single salt bridge contributes 3–5 kcal/mol to the free energy of folding of T4 lysozyme. *Biochemistry*, **29**, 2403–2408.
109. Serrano, L., Horovitz, A., Boaz, A., Bycroft, M. & Fersht, A. R. (1990). Estimating the contribution of engineered surface electrostatic interactions to protein stability by using double-mutant cycles. *Biochemistry*, **29**, 9343–9352.
110. Makhatadze, G. I., Loladze, V. V., Ermolenko, D. N., Chen, X. F. & Thomas, S. T. (2003). Contribution of surface salt bridges to protein stability: guidelines for protein engineering. *J. Mol. Biol.* **327**, 1135–1148.
111. Garcia-Mira, M. M. & Schmid, F. X. (2006). Key role of Coulombic interactions for the folding transition state of the cold shock protein. *J. Mol. Biol.* **364**, 458–468.
112. Robinson-Rechavi, M., Alibés, A. & Godzik, A. (2006). Contribution of electrostatic interactions, compactness and quaternary structure to protein thermostability: lessons from structural genomics of *Thermatoga maritima*. *J. Mol. Biol.* **356**, 547–557.
113. Stickle, D. F., Presta, L. G., Dill, K. A. & Rose, G. D. (1992). Hydrogen bonding in globular proteins. *J. Mol. Biol.* **226**, 1143–1159.
114. Huyghues-Despointes, B. M. P., Klingler, T. M. & Baldwin, R. L. (1995). Measuring the strength of side-chain hydrogen bonds in peptide helices: the Gln·Asp (*i*, *i*+4) interaction. *Biochemistry*, **34**, 13267–13271.
115. Muñoz, V. & Serrano, L. (1994). Elucidating the folding problem of helical peptides using empirical parameters. *Nature Struct. Biol.* **1**, 399–409.
116. Myers, J. K., Pace, C. N. & Scholtz, J. M. (1997). A direct comparison of helix propensity in proteins and peptides. *Proc. Natl Acad. Sci. USA*, **94**, 2833–2837.
117. Phillips, S. T., Piersanti, G. & Bartlett, P. A. (2005). Quantifying amino acid conformational preferences and side-chain–side-chain interactions in β-hairpins. *Proc. Natl Acad. Sci. USA*, **102**, 13737–13742.
118. Aurora, R. & Rose, G. D. (1998). Helix capping. *Protein Sci.* **7**, 21–38.

119. Creighton, T. E. (1988). Disulfide bonds and protein stability. *BioEssays*, **8**, 57–63.
120. Zdanowski, K. & Dadlez, M. (1999). Stability of the residual structure in unfolded BPTI in different conditions of temperature and solvent composition measured by disulfide kinetics and double mutant cycle analysis. *J. Mol. Biol.* **287**, 433–445.
121. Ramos, C. H. I. & Baldwin, R. L. (2002). Sulfate anion stabilization of native ribonuclease A both by anion binding and by the Hofmeister effect. *Protein Sci.* **11**, 1771–1778.
122. Sanchez-Ruiz, J. M. & Makhatadze, G. I. (2001). To charge or not to charge? *Trends Biotech.* **19**, 132–135.
123. Wunderlich, M., Martin, A. & Schmid, F. X. (2005). Stabilization of the cold shock protein CspB from *Bacillus subtilis* by evolutionary optimization of Coulombic interactions. *J. Mol. Biol.* **347**, 1063–1076.
124. Yao, M. & Bolen, D. W. (1995). How valid are denaturant-induced unfolding free energy measurements? Level of conformance to common assumptions over an extended range of ribonuclease A stability. *Biochemistry*, **34**, 3771–3781.
125. Whitten, S. T. & García-Moreno, B. (2000). pH dependence of stability of staphylococcal nuclease: evidence of substantial electrostatic interactions in the denatured state. *Biochemistry*, **39**, 14292–14304.

Edited by C. R. Matthews

(Received 30 January 2007; received in revised form 23 May 2007; accepted 25 May 2007)
Available online 2 June 2007