ELSEVIER

Biophysical Chemistry 85 (2000) 25–39

Biophysical Chemistry

www.elsevier.nl/locate/bpc

Heat capacity of hydrogen-bonded networks: an alternative view of protein folding thermodynamics

Alan Cooper*

Chemistry Department, Glasgow University, Glasgow G12 8QQ, Scotland, UK

Received 26 November 1999; received in revised form 7 March 2000; accepted 7 March 2000

Abstract

Large changes in heat capacity (ΔC_p) have long been regarded as the characteristic thermodynamic signature of hydrophobic interactions. However, similar effects arise quite generally in order–disorder transitions in homogeneous systems, particularly those comprising hydrogen-bonded networks, and this may have significance for our understanding of protein folding and other biomolecular processes. The positive ΔC_p associated with unfolding of globular proteins in water, thought to be due to hydrophobic interactions, is also typical of the values found for the melting of crystalline solids, where the effect is greatest for the melting of polar compounds, including pure water. This suggests an alternative model of protein folding based on the thermodynamics of phase transitions in hydrogen-bonded networks. Folded proteins may be viewed as islands of cooperatively-ordered hydrogen-bonded structure, floating in an aqueous network of less-well-ordered H-bonds in which the degree of hydrogen bonding decreases with increasing temperature. The enthalpy of melting of the protein consequently increases with temperature. A simple algebraic model, based on the overall number of protein and solvent hydrogen bonds in folded and unfolded states, shows how ΔC_p from this source could match the hydrophobic contribution. This confirms the growing view that the thermodynamics of protein folding, and other interactions in aqueous systems, are best described in terms of a mixture of polar and non-polar effects in which no one contribution is necessarily dominant.

Keywords: Protein folding thermodynamics; Heat capacity; Hydrogen bond network

* Tel.: +44-141-330-5278; fax: +44-141-330-2910; http://www.chem.gla.ac.uk/~alanc/alanc.html.
E-mail address: alanc@chem.gla.ac.uk (A. Cooper)

0301-4622/00/$ - see front matter
PII: S0301-4622(00)00136-8

1. Introduction

Non-covalent interactions involving biological and other macromolecules frequently involve large changes in heat capacity (C_p) of the system. The thermal unfolding of globular proteins in water is a classic example, where direct calorimetric (DSC) measurements, supported by more indirect methods, show that an overall heat capacity increment (positive ΔC_p) is an almost universal characteristic of the unfolding transition [1–8], see Fig. 1 for example. The heat capacity of the unfolded polypeptide is significantly greater than that of the more compact native conformation, which itself has a heat capacity comparable to that of an organic solid. Similar changes in C_p are seen in measurements of protein–protein interactions and, usually to a lesser extent, in the binding of smaller ligands to protein sites and other macromolecular processes.

Since heat capacity is the fundamental quantity from which other thermodynamic functions may be derived (particularly enthalpy and entropy), it is clearly important to clarify the source(s) of any apparently anomalous C_p behaviour. It is particularly important in biomolecular systems since ΔC_p effects might have significant functional consequences. For example, a large ΔC_p leads to significant temperature variations in ΔH and ΔS which, nevertheless, tend to compensate to give relative smaller changes in the more functionally significant Gibbs free energy change, ΔG, which is then less sensitive to temperature fluctuations [9] (Cooper et al., in preparation). Such variability makes it difficult to use enthalpy and/or entropy values alone for diagnosing specific components of non-covalent forces [10], though the 'thermodynamic buffering' that results may be of biological advantage. Entropy–enthalpy compensation, and the relatively high ΔC_p from which it stems, is often thought to be associated with the unusual properties of solvent water, but it is now known to be a rather general consequence of the multiplicity of weak, non-covalent interactions in macromolecular systems [11], of which liquid water is just one possible example. We would anticipate, nonetheless, that water should play the dominant role in biomolecular systems.

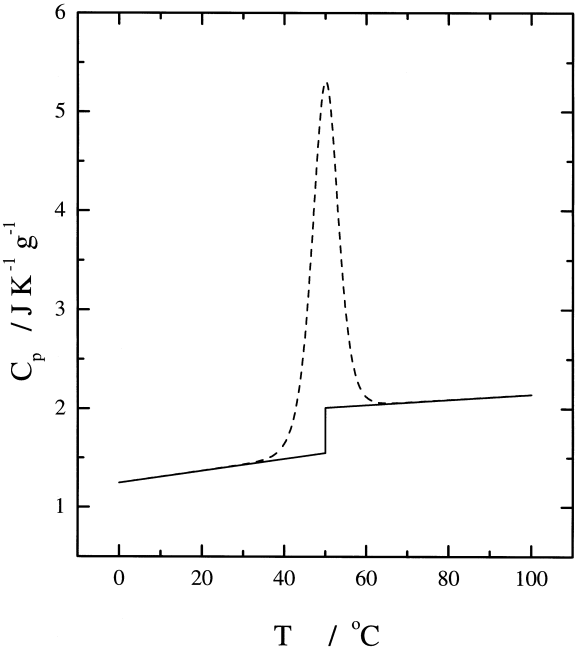

Fig. 1. Typical data for the heat capacity increment observed upon thermal unfolding of a globular protein in aqueous solution. Representative data for the excess specific heats (C_p) of folded and unfolded states (solid line) for a protein (lysozyme) unfolding cooperatively at 50°C are taken from [41] and references therein. The dotted line shows the overall heat energy uptake associated with the actual unfolding transition, for illustration, although it is the absolute heat capacities shown by the pre- and post-transition baselines that are of concern here. The heat capacity increment (ΔC_p) at the transition temperature is approximately 0.46 J K^{-1} g^{-1} in this case.

Speculation on the origin of ΔC_p effects is not new. Since Kauzmann [6,12], large changes in heat capacity have been taken as evidence for the involvement of hydrophobic interactions arising from specific properties of water. Experimental data on transfer of small molecules from non-aqueous to aqueous liquid environments show that such transfers are normally accompanied by a positive ΔC_p [13]. This is usually pictured at the molecular level as a consequence of the increase in hydrogen-bonded water structures that form around non-polar groups as they are accommodated into the liquid water lattice. The progressive 'melting' of such ice-like structures with increasing temperature soaks up thermal energy

and thereby accounts for the increased heat capacity. Accordingly, when a protein unfolds, hitherto buried hydrophobic residues become exposed to water, with resultant increase in C_p. Similar arguments can be applied to changes in side-chain environment during protein–protein and protein–ligand interactions. There has been considerable argument about whether this effect is sufficient to explain the magnitude of the effect seen in protein folding and other interactions, but it is generally believed that such changes in hydrophobic exposure are the source of ΔC_p effects and there is impressive correlation between measured ΔC_p and changes in accessible surface areas (ASA) in several systems [14]. Despite this, there are still some inconsistencies [6,15,16], and several authors have challenged the dominance of the role of hydrophobic interactions [17–19].

But changes in heat capacity are seen in many order–disorder transitions, and are not necessarily a sole property of hydrophobic interactions. Particularly relevant are the increases in C_p associated with the melting of polar solids which, as will be illustrated below, mimic in many ways the unfolding of proteins. Consequently, we will examine here an alternative view of ΔC_p effects based on phase changes in polar (H-bonded) networks, to see if such effects might be of relevance in biomolecular systems. These considerations will be based on known thermodynamic properties of pure compounds, together with a simple thermodynamic model for protein unfolding in an aqueous lattice. What will emerge is a picture illustrating how it is possible to rationalise, both qualitatively and quantitatively, some of the apparently anomalous heat capacity properties of biomolecules in water in terms of the overall solvent/macromolecular H-bonded network, without necessarily any need to resort to hydrophobic explanations. The purpose here, however, is not necessarily to supplant such explanations, but simply to show how additional factors may also contribute.

2. Heat capacities of solid–liquid transitions

An increase in heat capacity is a fundamental property of most order–disorder transitions in the condensed phase because of the increase in accessible degrees of freedom to molecules in the disordered state. If the disordered (liquid) state moreover allows thermal disruption of intermolecular interactions (bonds) which would otherwise be held in place by the cooperative nature of the crystal lattice, then such substances should show a further increase in heat capacity when going from solid to liquid phase. Empirical data confirm that the melting of most crystalline solids is accompanied by a positive ΔC_p, as shown in Fig. 2. This illustrates the general rule that melting of simple crystalline solids to a more disordered liquid state generally results in an increase in heat capacity of the substance. More particularly, as a general empirical rule, melting of polar materials involving intermolecular hydrogen bonding tends to give the highest values for ΔC_p, as is most evident for substances like water and ammonia where both the solid and the liquid phases are highly structured with extensive intermolecular hydrogen bonding. Liquids generally have a greater specific heat than solids under the same conditions. For example, compilations of experimental heat capacities [20] show that for pure organic compounds with elemental composition roughly similar to that of proteins, the heat capacities at 298 K fall into two discrete sets (Fig. 3a) depending on whether the substances are liquids or crystalline solids at that temperature. For CHNO compounds (up to C_5) for which data are available at 298 K, the mean heat capacity is 1.32 ($\pm$0.22) J K^{-1} g^{-1} for crystalline solids ($n = 118$; range 0.81–1.92) and 2.18 ($\pm$0.38) J K^{-1} g^{-1} for liquids ($n = 159$; range 1.53–4.16). Fig. 3b shows also the distribution of heat capacities of folded and unfolded proteins in solution, for comparison. This general trend is also illustrated by the empirical analysis of functional group contributions to the heat capacities of solid and liquid organic compounds [21], which shows that polar functional groups tend to contribute a higher C_p component to the liquid rather than the crystalline solid phase at the same temperature.

Heat capacity differences between solid and liquid phases may be rationalised in qualitative

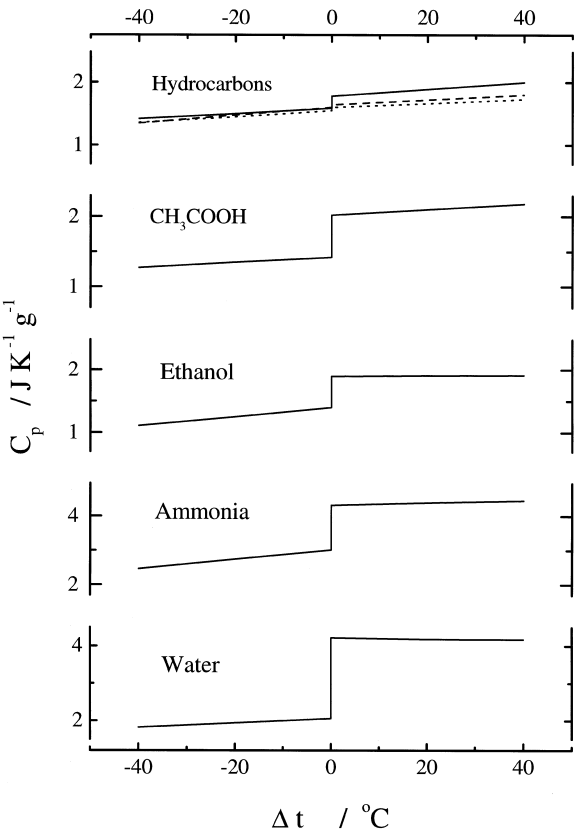

Fig. 2. Examples of absolute heat capacities for pure solid and liquid compounds as a function of temperature, plotted with respect to the normal melting point ($\Delta t = T - T_m$). Empirical data are taken from [52]. For simplicity, the heat capacity increments at the melting points ($\Delta t = 0$) omit the large C_p discontinuity associated with heat of melting at this point. The hydrocarbons (solid line: cyclohexane, $T_m = 6.7$°C, $\Delta C_p = 0.19$ J K^{-1} g^{-1}; dashed line: benzene, $T_m = 5.5$°C, $\Delta C_p = 0.04$ J K^{-1} g^{-1}; dotted line: naphthalene, $T_m = 81$°C, $\Delta C_p = 0.04$ J K^{-1} g^{-1}) generally show much smaller changes in heat capacity at the crystalline solid:liquid phase transition than the more polar compounds (water: $T_m = 0$°C, $\Delta C_p = 2.16$ J K^{-1} g^{-1}; ammonia: $T_m = -78$°C, $\Delta C_p = 1.31$ J K^{-1} g^{-1}; ethanol: $T_m = -130$°C, $\Delta C_p = 0.67$ J K^{-1} g^{-1}; ethanoic (acetic) acid: $T_m = 16.5$°C, $\Delta C_p = 0.59$ J K^{-1} g^{-1}).

terms by observing that the crystalline phase of most polar compounds comprises a cooperatively hydrogen bonded network or lattice where thermal breakage of individual bonds is energetically and structurally unfavourable in the solid state. This means that heat absorption is confined mainly to the vibrational degrees of freedom allowed by this ordered lattice and consequently C_p is low. This contrasts with the liquid state where, although residual H-bond/lattice structure may persist (as seen to greatest effect in liquid water), intermolecular bonds have much more freedom to bend or break with increasing temperature. It follows, therefore, that additional thermal energy

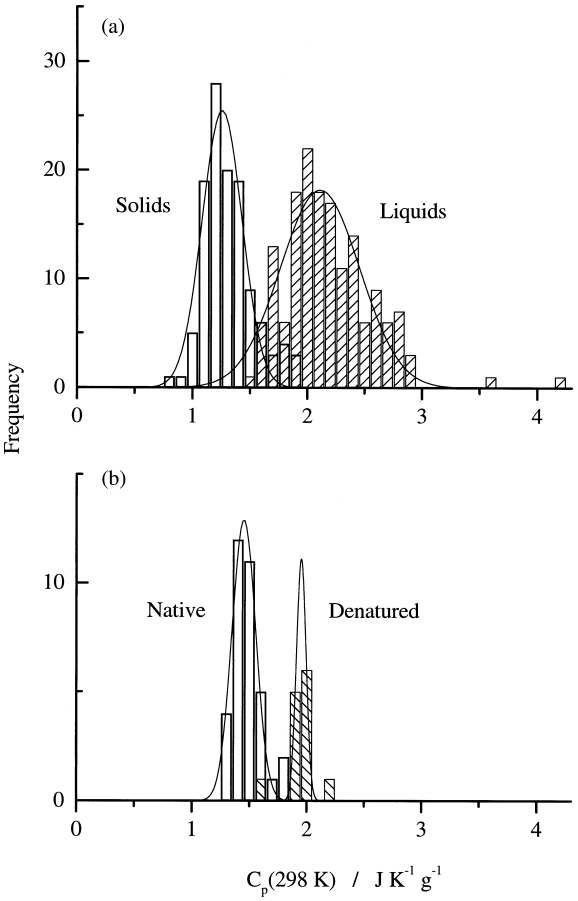

Fig. 3. (a) Distribution of specific heats (heat capacity, C_p) at 298 K for pure organic solids and liquids. Experimental data are taken from [20] for compounds up to and including C5 with elemental composition CHO, CHN or CHNO for which data are available (118 crystalline solids and 159 pure liquids). The mean values (±S.D.) are 1.32 (±0.22) J K^{-1} g^{-1} and 2.18 (±0.38) J K^{-1} g^{-1} for solids and liquids, respectively. The curves show Gaussian fits to the individual distributions. (b) C_p values for native ($n = 35$) and denatured ($n = 13$) proteins in aqueous solution at 298–300 K. Data are taken from the compilation in [41] with mean values of 1.49 (±0.13) J K^{-1} g^{-1} and 1.95 (±0.15) J K^{-1} g^{-1} for native and denatured states, respectively.

may be taken up by the polar liquid as intermolecular H-bonds are broken, in addition to the translational and rotational degrees of freedom further allowed to the molecules in the liquid state. Consequently it takes more thermal energy to bring about a rise in temperature of the liquid than the solid, and C_p is higher.

3. Proteins as parts of H-bonded networks

The heat capacity effects described above for pure substances show striking similarity with changes seen upon thermal unfolding of proteins (Figs. 1–3). The apparent excess heat capacities of proteins in solution can now be measured quite accurately by calorimetric methods [1–8,22]. For native globular proteins in water the heat capacity shows a slight temperature dependence, with typical values of approximately 1.5 J K^{-1} g^{-1} at room temperature, comparable to organic solids of similar elemental composition (Fig. 3). Unfolded polypeptides in water have a higher heat capacity (approx. 1.9 J K^{-1} g^{-1}), and the overall temperature dependence of C_p (including the unfolding transition) as typically observed in a DSC experiment is illustrated in Fig. 1. [Unlike regular crystalline solids, the microscopic size of a single protein molecule means that the 'melting' transition does not show the discontinuity in heat capacity usually seen for first-order phase transitions of macroscopic systems, but rather the broader endotherm — shown by the dotted line in Fig. 1 — representing the 'latent heat of fusion' or thermal energy required to undergo unfolding. However, this simply reflects the finite size of the cooperative unit (i.e. the protein molecule) involved in the transition [3,7] and has no bearing on the more general properties under consideration here.]

These similarities — both qualitative and quantitative — between the melting of pure compounds and the unfolding of proteins in water may be simply coincidental. However, on reflection, there may be closer parallels than hitherto considered. Globular proteins in their native state are generally compact cooperative structures, and their thermal unfolding may be regarded in some ways as analogous to the melting of a small solid. Other thermodynamic properties of native proteins such as packing density, compressibility, thermal expansion coefficient, etc., are similar to bulk organic solids [19,23–25], and recent structural analysis [26] has shown how they may be best regarded as surface-molten solids. A folded protein is a cooperative structure in which most, if not all, polypeptide backbone and sidechain hydrogen bonds are satisfied, either internally by H-bonding to other groups or externally by solvation. Extensive structural studies have shown that unpaired H-bond donors or acceptors are rare inside native proteins [27,28], as might be expected from the energetic cost of breaking a hydrogen bond without compensatory solvation or other interactions. The interior of a globular protein, although often pictured as a non-polar environment, more realistically comprises an extended network of hydrogen-bonded interactions between peptide groups and polar sidechains, irregular but nonetheless conformationally specific, within which non-polar residues may be accommodated. Indeed, it is this cooperative interplay between the packing of bulky sidechains and the satisfaction of H-bonding requirements that determines the native fold. And the cooperative nature of the fold means that, although there may be small changes in dynamics and surface details, the majority of these structural features remain intact in the folded state regardless of variations in temperature, pH, or the presence of denaturants, until the system becomes globally unstable and a cooperative transition to some other state is induced.

In distinct contrast to the cooperative native fold, the unfolded protein is a much less ordered structure: conformationally heterogeneous and dynamic, with groups exposed predominantly to solvent water. Many of the protein–protein H-bonds that existed in the folded structure may be replaced by protein–water interactions in the unfolded state. However, the extent of hydrogen bonding — both peptide–water in the unfolded state and in the aqueous solvent itself — will vary with temperature (and denaturants?) in a way that should be reflected in the temperature dependence of the enthalpy of unfolding and other

parameters. (The molecular mechanics of protein–water hydrogen bonds and the effects of local water bridges on polypeptide conformations have been examined recently [29], but here we wish to focus on the more general thermodynamic effects on protein stability.) The situation is illustrated schematically in Fig. 4. At low temperatures, protein–protein hydrogen bonds within the folded structure are largely replaced by hydrogen bonds to water molecules in the unfolded state, so there is little or no change in overall numbers of bonds. At higher temperatures, however, much less H-bonding network exists in the solvent and unfolded polypeptide states — although it persists in the cooperatively folded state. Consequently there is an overall breakage of hydrogen bonds upon unfolding at higher temperatures, and the enthalpy of unfolding becomes progressively more endothermic with increase in temperature.

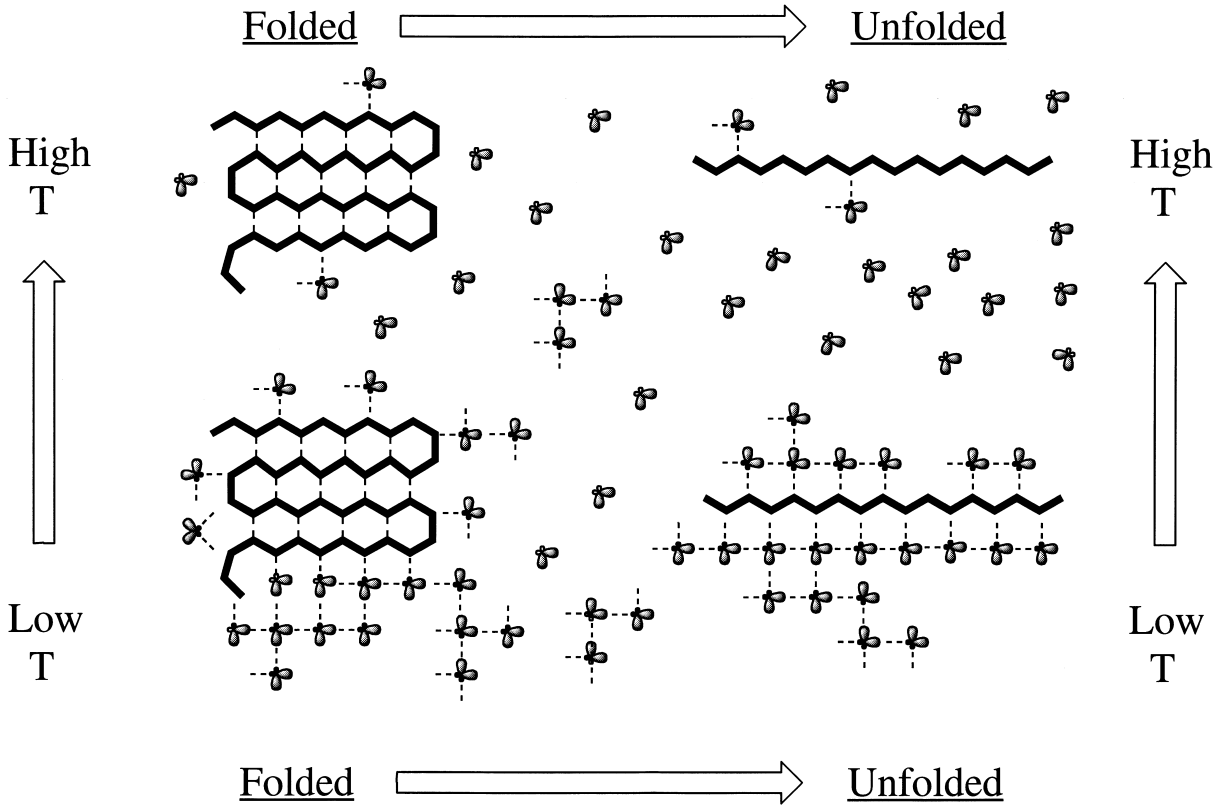

Fig. 4. Cartoon depicting the extent of protein/water hydrogen bonding in the folded and unfolded states. No semblance of structural realism is intended here. The bold lines represent stylistically the protein polypeptide, with water molecules swarming around. Hydrogen bonds are indicated by dotted lines. The cooperatively folded structure will have elements of secondary and tertiary structure with an internal hydrogen bonded network that remains relatively intact regardless of temperature. In contrast, the unfolded polypeptide, and the water that surrounds and hydrates it, will contain more transient hydrogen bonded networks that decrease in extent with increase in temperature. (This unfolded polypeptide, though sketched here in an extended conformation, will in reality exist as a dynamic ensemble of conformational substates that approaches the hypothetical random coil state only under extreme conditions.) When the protein unfolds at low temperatures, most, if not all internal hydrogen bonds might be replaced by hydrogen bonds to water, and the net enthalpy contribution from this source might be close to zero. However, for unfolding at higher temperatures, the cooperative nature of the protein fold means that the same number of internal bonds are broken, but relatively fewer of them might be replaced by interaction with water. Consequently, unfolding at higher temperature will be more endothermic.

All this takes place, normally, within a solvent environment which itself consists of an extensive, albeit dynamic, network of hydrogen bonded interactions between water molecules. There have been many theories of liquid water, but… '*The present consensus seems to be that liquid water is a macroscopic network of molecules connected by frequent but transient hydrogen bonds, which allow unbonded neighbours to occur in numbers that vary with temperature and pressure*.' [30]. (As pointed out by Dill [6], this 'melting' of the hydrogen-bonded water lattice at higher temperatures might best be viewed as a bending rather than breakage of discrete bonds, but this does not affect the overall picture presented here.) Consequently, both the protein molecule and the water environment around it comprise extensively hydrogen-bonded networks. For the liquid water and the unfolded polypeptide within it, this network is fluxional, with ephemeral regions of order and disorder in perpetual dynamic interchange, at rates and levels that depend on temperature and other parameters. In comparison, folded proteins may be viewed as islands of cooperatively-ordered hydrogen-bonded structure, floating in an aqueous network of less-well-ordered H-bonds in which the degree of hydrogen bonding decreases with increasing temperature.

One consequence of this picture is that the role of solvent water in the overall thermodynamics of hydrogen bonding in such systems should now become more temperature dependent. Conventionally, hydrogen bonding in biomolecular systems has been considered 'thermodynamically neutral', since it was generally perceived that in a typical unfolding or ligand-dissociation process, H-bonds between protein (or other) groups would be replaced by H-bonds to water in the overall scheme, and no net change in total number of hydrogen bonds would result. This conclusion follows naturally from the known high solubility of polar compounds in water, supported in more detail by classic model compound studies [6,19,31,32] which seem to confirm that little or no thermodynamic advantage to the formation of hydrogen bonds between groups when water was present — at least in terms of free energy (ΔG). In this picture, hydrogen bonds are clearly crucial in controlling folding and stereochemical specificity, but do not appear to contribute in any major way to the overall thermodynamic stability. However, such model compound studies have been concerned mainly with ΔG effects at just one temperature (usually 25°C), so heat capacity effects have not been examined in detail. Moreover, the effects of cooperativity and temperature on the overall H-bond balance sheet are not accounted for in such models. In particular, if the propensity for water hydrogen bonding to solvated groups decreases at higher temperatures, then the overall neutrality of hydrogen bonding must be called into question. In the context of protein unfolding, as described above, this means that the total number of hydrogen bonds broken in the protein–water system during unfolding will rise as the temperature of the transition increases. Consequently, at least enthalpically, hydrogen bonding will not be neutral over an extended temperature range. Indeed, we would expect the endothermic ΔH arising from hydrogen bond breaking to increase with temperature, giving a positive contribution to the ΔC_p of unfolding (see Fig. 4).

4. Protein unfolding: a hydrogen-bonded lattice model

These largely qualitative observations may be reinforced by consideration of a simple algebraic model taking account of the overall numbers of hydrogen bonds, including solvent, in both folded and unfolded states at different temperatures. For simplicity, and in order to illustrate the main points, this model will be based on the following assumptions:

1. The native protein (N) is a compact, cooperatively-folded structure in which most, if not all, internal hydrogen bonding interactions are satisfied. The cooperative nature of the folding means that these hydrogen bonds remain essentially intact at all temperatures, until the protein unfolds. (This is essentially the '2-state' hypothesis upon which most analysis of protein folding is based.)

2. The unfolded polypeptide (U) is an irregular, flexible structure (or ensemble of dynamically interchanging structures), predominantly exposed to solvent. The non-cooperative nature of this structure means that interactions are kinetically transient and temperature dependent. In particular, the extent of hydrogen bonding between protein (peptide and sidechain) groups and water will decrease with rise in temperature.
3. The solvent water (w) comprises a similarly irregular hydrogen-bonded lattice, in which the extent of intermolecular H-bonding also decreases with increasing temperature.

Clearly these assumptions are simplistic, particularly in relation to the structure and interactions in liquid water, but they will allow us to proceed to illustrate the salient features. Much more realistic details may be incorporated into the model later, if required, once its relevance has been established.

The model will examine what contribution hydrogen bonds might make, over and above hydrophobic and other non-covalent interactions that have already been extensively discussed by others. We will concentrate here on counting the total number of hydrogen bonds in the system under various conditions, and for this it is necessary to define a series of terms. Imagine an isolated protein molecule in an appropriate amount of solvent water, and let:

n = the total number of amino acid residues in the polypeptide chain;
n_w = the total number of water molecules (per polypeptide) in the system;
f_{pp} = the fraction of (internal) protein–protein H-bonds made in the natively-folded (N) state ($f_{pp} = 1$ in the simplest case, ignoring unmade bonds);
f_{ww} = the fractional probability of formation of water–water H-bonds in the bulk solvent;
f_{pw} = the fractional probability that any exposed protein donor/acceptor group may be H-bonded to water in the unfolded (U) state;
h_{pp} = the (average) enthalpy of formation of a protein–protein H-bond, assumed all identical for simplicity;
h_{ww} = the enthalpy of formation of a water–water H-bond;
h_{pw} = the (average) enthalpy of formation of a protein–water H-bond, assumed all identical for simplicity;
H_N = total enthalpy of the native protein system (including waters); and
H_U = total enthalpy of the unfolded protein system (including waters).

Similar nomenclature will apply for other thermodynamic parameters: entropy (s_{ww}, s_{wp}, s_{pp}, S_N, S_U); and free energy (g_{ww}, g_{wp}, g_{pp}, G_N, G_U).

Imagine a polypeptide chain that can fold cooperatively into a compact globular structure and in which donor and acceptor sites may align (more or less) perfectly to form bonds. Imagine also that this exists in a sea of solvent (water) molecules that themselves also carry donor and acceptor sites that may form bonds with each other or with exposed sites on the polypeptide or protein (see Fig. 4). Since we are interested in changes in thermodynamic parameters during unfolding, assume a scenario in which there exist only two possible states: the compact 'native' fold (N) and the flexible, extended, 'unfolded' state (U). (This does not necessarily imply that only these two states are available, but simply that these represent two extreme possibilities. The nature of cooperative 2-state behaviour has been discussed by several authors, see [33,34] for example, and even a continuum of states can exhibit apparent 2-state properties under some circumstances [35].)

In order to estimate the hydrogen bonding contribution to the overall thermodynamics of unfolding, it is necessary simply to count the total number of hydrogen bonds (including solvent bonds) in each state. A fraction (f_{pp}) of donor–acceptor groups in the native protein (N) will form protein–protein (p–p) bonds. In the unfolded protein (U) a fraction (f_{pw}) of such groups will instead form hydrogen bonds to water. Both N and U systems will also contain 'free' water molecules that will themselves experience

transient hydrogen bonding with other molecules, with time-averaged fractional water–water hydrogen bonding f_{ww}. Simply counting hydrogen bonds, including both chain and water molecules and bearing in mind that each residue (peptide) contains two potential hydrogen bond sites and that solvent bonds are shared, gives the total enthalpy of the N state (assuming only p–p and w–w bonds exist, and ignoring sidechain hydrogen bonds for simplicity):

$$H_N = n \cdot f_{pp} \cdot h_{pp} + n_w \cdot f_{ww} \cdot h_{ww}/2$$

and for the U state:

$$H_U = 2n \cdot f_{pw} \cdot h_{pw} + (n_w - 2n) \cdot f_{ww} \cdot h_{ww}/2$$

giving, for the enthalpy of unfolding (hydrogen bond contribution only):

$$\Delta H_{unf,h} = H_U - H_N = n \cdot (2 \cdot f_{pw} \cdot h_{pw} - f_{ww} \cdot h_{ww} - f_{pp} \cdot h_{pp}) \quad (1)$$

with similar expressions for the hydrogen bond contributions to unfolding entropy and free energy. (The subscript 'h' indicates that this represents only hydrogen bond contributions. Other terms may be added later if required.)

This expression [Eq. (1)] will allow qualitative discussion, and might even form the basis for simple numerical estimates. For example, and to check for consistency, consider some particular cases.

Firstly, imagine the situation if all hydrogen bonds are identical and of equal occupancy no matter where they occur, then $h_{ww} = h_{pw} = h_{pp}$, $f_{ww} = f_{pw} = f_{pp}$, and $\Delta H_{unf,h} = 0$ as expected. In this case, which corresponds roughly to the conventional thermodynamically neutral view of hydrogen bond effects in water, there would be no enthalpic contribution (stabilising or otherwise) to folding. Similar conclusions would be drawn about hydrogen bond contributions to ΔS_{unf} and ΔG_{unf}. In such a case, all the thermodynamics would be determined by the other interactions (hydrophobic, dispersion, etc.) that have been omitted here.

A slightly more plausible view might propose (for simplicity) that all hydrogen bonds have the same enthalpies, $h_{ww} = h_{pw} = h_{pp} \equiv h$, but (potentially) different fractional occupancies. Although it is clearly unrealistic to assume that all H-bond enthalpies are the same, regardless of their chemical nature, this assumption does allow formulation of a basic model that may be refined later as necessary. In this case:

$$\Delta H_{unf,h} = nh(2 \cdot f_{pw} - f_{ww} - f_{pp})$$

showing that a finite enthalpy of unfolding (positive or negative) can arise even if all hydrogen bonds have the same heat of formation, provided the overall extent of hydrogen bonding changes during the unfolding process.

This example forms the basis for a tentative working model in which we might assume for simplicity, or because we lack any more detailed knowledge, that:

1. all internal protein hydrogen bonds are made in the native state: $f_{pp} = 1$;
2. all hydrogen bond enthalpies are equal: $h_{ww} = h_{pw} = h_{pp} \equiv h$; and
3. protein–water and water–water bonds are indistinguishable: $f_{pw} = f_{ww} \equiv f$.

This reduces Eq. (1) to a seductively simple expression for the unfolding enthalpy:

$$\Delta H_{unf,h} = -nh(1 - f) \quad (2)$$

where h is the heat of formation of a hydrogen bond (of whatever kind) and f is the fractional probability of any water molecule or exposed protein group forming a hydrogen bond with another water molecule.

Certain consequences follow from this model, as expressed by Eq. (2):

1. $\Delta H_{unf,h}$ is normally *endothermic* (positive) since the heat of hydrogen bond formation (h) must be negative (exothermic), and $f \leq 1$.
2. Since, intuitively, f will decrease with increasing temperature, $\Delta H_{unf,h}$ also increases (becoming more endothermic) with increasing

temperature. Consequently, $\Delta C_{p,unf}$ is *positive*.

3. $\Delta H_{unf,h}$ might reach a *hypothetical maximum* ($\Delta H_{unf,h,max} = -nh$) when $f = 0$, presumably at some temperature sufficiently high above 100°C that water molecules (in the liquid state) might be viewed to have lost all hydrogen bond interactions with their neighbours.
4. $\Delta H_{unf,h}$ might reach a *hypothetical minimum* ($\Delta H_{unf,h,min} = 0$) at some sufficiently low (but possibly unattainable) temperature when f approaches unity.
5. Exothermic unfolding at low temperatures (cold denaturation [36,37]), though not explicitly a prediction of this primitive model [Eq. (2)], would result if not all p–p bonds were made in the native state ($f_{pp} < 1$). In such instances Eq. (2) would become: $\Delta H_{unf,h} = -nh(f_{pp} - f)$ so that the enthalpy of unfolding would change sign at temperatures below which the fractional hydrogen bonding in (supercooled?) liquid water exceeds that in the folded protein. Since 100% of possible H-bonding is not seen even in the best folded examples [27,28], this is not implausible.

All these predictions match remarkably well with known properties of proteins [1–7,22], despite the relative simplicity of the underlying assumptions. More detailed versions of this model might adopt the more realistic view (e.g. [19,38]) that not all hydrogen bonds are the same in Eq. (1) and use a variety of enthalpies (h) and fractional probabilities (f). For example, if formation of protein–water hydrogen bonds were to be more exothermic than water–water or protein–protein bonds ($h_{pw} < h_{ww}, h_{pp}$) then protein unfolding would be exothermic even without change in f. The unfolding enthalpy could thus be made either endo- or exothermic at will by appropriate choice of parameters. Unfortunately there is not yet sufficient detailed empirical information about the individual hydrogen bond strengths in such circumstances [19] to merit taking this approach further here.

Similar considerations will also apply to the entropic contribution from hydrogen bonding. Again taking the simplest basic assumptions as above, the hydrogen bond component of the entropy of unfolding might be written:

$$\Delta S_{unf,h} = -ns(1 - f)$$

where s is the entropy of formation of any hydrogen bond, expected to be negative because of the loss of motional degrees of freedom that this implies. As a result, $\Delta S_{unf,h}$ would be expected to make a positive contribution to unfolding, again increasingly positive with increase in temperature as f decreases.

The relative stability of N vs. U depends of course on the balance of enthalpic and entropic contributions as expressed in the overall Gibbs free energy change, which for the hydrogen bond contribution in this model would be written:

$$\begin{aligned}\Delta G_{unf,h} &= \Delta H_{unf,h} - T \cdot \Delta S_{unf,h} \\ &= -n(1 - f)(h - T \cdot s)\end{aligned}$$

For completeness at this stage, recognising that we have ignored or omitted many other possible contributions to the overall unfolding process, we must add the change in configurational entropy of the polypeptide chain upon unfolding (ΔS_{config}; this is of course the major thermodynamic barrier to folding that all other interactions must overcome if folding is to be possible), together with a contribution (ΔG_{other}) that takes account of all other interactions (hydrophobic, electrostatic, van der Waals, etc.) so far neglected. This gives an overall expression for the unfolding free energy of the form:

$$\begin{aligned}\Delta G_{unf} = &-n(1 - f)(h - T \cdot s) - T \cdot \Delta S_{config} \\ &+ \Delta G_{other}\end{aligned}$$

where the first term on the right is the hydrogen bond model contribution developed here, T is the absolute temperature, ΔS_{config} is the change (increase) in configurational entropy of the polypeptide chain upon unfolding, and ΔG_{other} covers all other contributions to the unfolding free energy.

Certain questions remain:

1. What is meant by 'h' and 's' in this context? We have defined these terms as the enthalpy

(h) and entropy (s) of formation of a hydrogen bond in the protein/solvent lattice, without much consideration of what they might represent in real terms or what values they might take. As a first approximation we might imagine these to be similar to the enthalpy and entropy of hydrogen bond formation between groups in vacuo (or vapour phase), and we might certainly use such numbers as an upper limit. Enthalpies (h) based on such estimates would be in the range -12 to -25 kJ mol^{-1} (-3 to -6 kcal mol^{-1}), for example [19,39]. But this is probably an overestimate. More reasonably, in the context of the current model, h and s might be considered to be the enthalpy and entropy change for a (possibly hypothetical) process going from a freely rotating group in a polar cavity to a fixed, hydrogen bond orientation. As a result, s will probably be negative due to loss of rotational degrees of freedom, and h will be the (negative) enthalpy difference between a rotationally averaged-dipole in a polar cavity and the most favourable hydrogen bond orientation. In either case, the absolute magnitudes are probably less than would be anticipated from in vacuo measurements.

2. How might we estimate f and its variation with temperature?

Inasmuch as we are assuming that p–w bonds are very similar to w–w bonds, and inasmuch as many of the thermodynamic properties of liquid water are themselves a manifestation of hydrogen bond interactions, it may be reasonable to estimate $f(T)$ from the known properties of water ([39], for example). Following the same logic as above, considering liquid water alone in the absence of protein, and assigning the bulk of the thermodynamic properties to hydrogen bonding, the total enthalpy of n_w water molecules in the liquid phase would be:

$$H_w = n_w \cdot f_{ww} \cdot h_{ww}$$

with molar heat capacity: $C_{p,water} = (1/n_w) \cdot dH_w/dT = h_{ww} \cdot df_{ww}/dT$, giving: $df_{ww}/dT = C_{p,water}/h_{ww}$.

Incorporating this into the primitive model [Eq. (2)] gives, for the molar heat capacity increment upon unfolding:

$$\Delta C_{p,unf} = d\Delta H_{unf}/dT = nh \cdot df/dT \approx n \cdot C_{p,water}$$

where $C_{p,water}$ is that fraction of the molar heat capacity of liquid water that might be ascribed to thermal disruption of the hydrogen bond network in the liquid. Of course, not all of the heat capacity of water (totalling 4.18 J K^{-1} g^{-1}, or 75 J K^{-1} mol^{-1}, at room temperature) may be attributed to the effects of hydrogen bonding. But the most that might be ascribed to molecular motions in the equipartition limit would be $3R$, or approximately 25 J K^{-1} mol^{-1}, assuming three translational and three rotational degrees of freedom (where R is the gas constant, and ignoring molecular vibrations that will be suppressed by quantum effects at normal temperatures). The remainder must be due to inter-molecular effects, primarily hydrogen bonding in this case. This would suggest an upper limit to the hydrogen bond contribution to the heat capacity increment on unfolding per residue of order $C_{p,water}$, that is approximately 50 J K^{-1} mol^{-1} at room temperature. Given the extreme crudity of the model, and the naivety of some of the underlying approximations, this is surprisingly good numerical agreement with observation. Experimental heat capacity increments for a range of proteins [22,40,41] lie in the range 40–80 J K^{-1} mol^{-1} per residue at room temperature, falling to lower values (20–50 J K^{-1} mol^{-1}) at higher temperatures in a way that the model does not (yet) predict — though this could easily be explained away by a reduction in df/dT or a temperature-dependent change in bonding in the native state. Nonetheless, this estimated upper limit per residue for ΔC_p of approximately 50 J K^{-1} mol^{-1}, corresponding to approximately 0.45 J K^{-1} g^{-1} for a mean residue weight of 110 Da, is not inconsistent with the observed spread of data at 25°C (Fig. 3). For the proteins for which data are currently available [41] the mean increase in C_p upon unfolding at 25°C is 0.48 ($\pm$0.14) J K^{-1} g^{-1}.

Further numerical estimates are equally promising. The enthalpy of the solid → liquid (ice → water) melting transition varies from 6 kJ mol^{-1} at 0°C to approximately 10 kJ mol^{-1} at 100°C (by extrapolation using experimental heat capacity data), suggesting an estimate for h of

approximately -5 kJ mol^{-1} if all this comes from breaking of hydrogen bonds (two per molecule) in the condensed phase. Similar orders of magnitude may be obtained from estimates of the fractional hydrogen bonding in liquid water at different temperatures [39]. Consequently we would anticipate a maximum ΔH_{unf} $(f \rightarrow 0)$ of order $+5$ kJ mol^{-1} per residue in this model. Experimentally [40] the maximal/plateau values of specific unfolding enthalpies are found to cluster approximately $+5.5$ kJ mol^{-1}. (Earlier work suggesting more exact convergence on common values at high temperatures, which sparked considerable speculation [2,42–44], has not been substantiated by data from a larger set of trial proteins [40].)

5. Discussion

Unfolding of a protein requires that specific hydrogen bonds within the native fold be replaced, at least in part, by less specific hydrogen bonds to solvent. The cooperative nature of protein folding implies that most, if not all hydrogen bonds are formed in the native state, but the number of H-bonds (primarily to solvent) in the unfolded state will depend on conditions such as temperature and denaturant concentrations. In particular, with increase in temperature, the number of peptide–solvent hydrogen bonds will decrease, giving an increasingly more endothermic contribution (positive ΔC_p) to the unfolding thermodynamics. Although one might query the numerical assumptions and approximations involved, it is clear that the thermodynamics of melting of hydrogen-bonded networks described here is at least qualitatively similar to the experimental behaviour for unfolding of proteins in water. And crude numerical estimates suggest that the model might be quantitatively significant as well. Using a model in which folded proteins are viewed as islands of crystallinity floating in a sea of less-well-ordered hydrogen bonds we have been able to account for many of the basic features of unfolding thermodynamics in terms of hydrogen bonding alone, without recourse to other effects.

But this contrasts with the more conventional view of protein folding, which attributes most of the thermodynamic effects to changes in exposure of non-polar sidechain groups during unfolding. Indeed, statistical correlation analyses which demonstrate a relationship between ΔC_p and buried surface areas (ASA) tend to indicate a negative contribution of peptide hydrogen bonding to unfolding heat capacity changes [14], in direct contrast to the positive value suggested here. Although the correlations look quite convincing, Rose and Wolfenden, and others, [19] have pointed out a potential difficulty with such analyses where… *'if proteins bury a well-behaved and constant fraction of their apolar surface and also have a constant heat capacity, then any strong correlation between these properties need not be causal..'* [19]. The universality of the ASA/ΔC_p effect has been challenged in recent work [45,46] in which ΔC_p values obtained experimentally do not necessarily tally with what might have been expected on the basis of changes in non-polar surface areas alone. In any case, the experimental situation is so under-determined, with broadly similar data from a relatively small set of proteins for which both detailed structural and calorimetric data are available, that it is feasible that multi-variate analysis may not necessarily yield a unique correlation. This possibility is currently under investigation, but it is feasible that in attempting to ascribe the majority of the ΔC_p effect to non-polar groups, one might force the analysis to over-compensate with other contributions that do not necessarily reflect physical reality. Moreover, the various small model compound systems — both solid and liquid — used in much previous work may not properly reflect the folded protein environment. Estimates based on solubilities and heats of solution of cyclic dipeptides [38] suggest a negative ΔC_p for amide–amide hydrogen bonds but a positive ΔC_p for amide–hydroxyl bonds. Such differences are hard to reconcile theoretically, and the experimental situation may be confounded by specific crystal packing effects that are difficult to separate from individual group contributions [38]. This illustrates some of the profound difficulties encountered when attempting to tease out the separate contributions to

interaction thermodynamics in such heterogeneous systems.

This work has shown how it is possible that 'partial melting' of hydrogen-bonded lattices might result in thermodynamic characteristics similar to those changes conventionally associated with hydrophobic group exposure. Consequently, in most cases involving unfolding of globular proteins, for example, it will be difficult to estimate which (if any) of these two interpretations is correct. One test would be to examine thermal transitions in polypeptides that do not contain significant buried hydrophobic groups or which do not involve significant changes in non-polar ASA. In such instances we would anticipate a finite positive ΔC_p in the unfolding transition, even in the apparent absence of hydrophobic effects. Two recent examples show just this effect [47,48]. A 16-residue peptide that forms a β-hairpin involving anti-parallel β-sheet in water unfolds cooperatively with a large $\Delta C_p = 1400$ J K^{-1} mol^{-1} (approx. 0.8 J K^{-1} g^{-1}) at 298 K [47]. A similar effect, with $\Delta C_p = 460$ J K^{-1} mol^{-1} (approx. 0.15 J K^{-1} g^{-1}), has been reported for the unfolding of an α-helical dicyclic 29-residue peptide [48]. These values span the range seen for unfolding of much larger globular proteins [22,40,41]. Although both sets of authors [47,48] attempt to interpret their data in hydrophobic terms, it might seem more pragmatic to associate the effects more directly with the more obvious hydrogen bonding changes. Similar ΔC_p effects are seen for unfolding of large fibrous proteins such as collagen, myosin and paramyosin [22,49], further showing that significant ΔC_p effects can occur even in the absence of an identifiable core of hydrophobic residues. Recent work has shown that single-layer β-sheets can fold without a hydrophobic core [50], and members of the cellular retinoic acid binding protein (CRABP) family have an unusually small hydrophobic core despite their very stable fold [51]. Although ΔC_p data are not yet available, such proteins and synthetic peptides should provide an interesting test of the proposals presented here.

The model described here might be applicable in other areas of protein folding and interaction. For example, the effects of increasing concentrations of denaturants such as urea or guanidinium hydrochloride might be viewed, at least in part, as analogous to an increase in temperature, since such small hydrophilic molecules are likely to disrupt the H-bonding lattice in the solvent and unfolded polypeptide state without markedly affecting the folded protein. It should be possible to extend these ideas to other systems such as protein–protein or protein–nucleic acid complexes where extensive hydrogen bonding and solvent interface effects may be significant. And the model may also rationalise the thermodynamics of non-specific aggregation of unfolded or mis-folded proteins frequently observed especially at higher temperatures.

References

[1] W.M. Jackson, J.F. Brandts, Thermodynamics of protein denaturation. Calorimetric study of the reversible denaturation of chymotrypsinogen and conclusions regarding the accuracy of the two-state approximation, Biochemistry 9 (1970) 2294–2301.
[2] P.L. Privalov, Stability of proteins: small globular proteins, Adv. Protein Chem. 33 (1979) 167–241.
[3] J.M. Sturtevant, Biochemical applications of differential scanning calorimetry, Ann. Rev. Phys. Chem. 38 (1987) 463–488.
[4] P.L. Privalov, S.J. Gill, Stability of protein structure and hydrophobic interactions, Adv. Protein Chem. 39 (1988) 191–234.
[5] P.L. Privalov, G.I. Makhatadze, Heat capacity of proteins. II. Partial molar heat capacity of the unfolded polypeptide chain of proteins, J. Mol. Biol. 213 (1990) 385–391.
[6] K.A. Dill, Dominant forces in protein folding, Biochemistry 29 (1990) 7133–7155.
[7] A. Cooper, Thermodynamics of protein folding and stability, in: G. Allen (Ed.), Protein: A Comprehensive Treatise, vol. 2, JAI Press Inc, Stamford, CT, 1999, pp. 217–270.
[8] M. Häckel, H-J. Hinz, G.R. Hedwig, A new set of peptide-based group heat capacities for use in protein stability calculations, J. Mol. Biol. 291 (1999) 197–213.
[9] D. McPhail, A. Cooper, Thermodynamics and kinetics of dissociation of ligand-induced dimers of vancomycin antibiotics, J. Chem. Soc. Faraday Trans. 93 (1997) 2283–2289.
[10] A. Cooper, Thermodynamic analysis of biomolecular interactions, Curr. Opin. Chem. Biol. 3 (1999) 557–563.
[11] J.D. Dunitz, Win some, lose some: enthalpy–entropy compensation in weak intermolecular interactions, Chem. Biol. 2 (1995) 709–712.

[12] W. Kauzmann, Some factors in the interpretation of protein denaturation, Adv. Protein Chem. 14 (1959) 1–63.

[13] C. Tanford, The Hydrophobic Effect: Formation of Micelles and Biological Membranes, 2nd ed, Wiley, New York, Chichester, 1980.

[14] R.S. Spolar, J.R. Livingstone, M.T. Record, Jr., Use of liquid hydrocarbon and amide transfer data to estimate contributions to thermodynamic functions of protein folding from the removal of nonpolar and polar surface from water, Biochemistry 31 (1992) 3947–3955.

[15] T. Lazaridis, G. Archontis, M. Karplus, Enthalpic contributions to protein stability: insights from atom-based calculations and statistical mechanics, Adv. Protein Chem. 47 (1995) 231–306.

[16] B. Hallerbach, H.-J. Hinz, Protein heat capacity: inconsistencies in the current view of cold denaturation, Biophys. Chem. 76 (1999) 219–227.

[17] C.N. Pace, B.N. Shirley, M. McNutt, K. Gajiwala, Forces contributing to the conformational stability of proteins, FASEB J. 10 (1996) 75–83.

[18] J.K. Myers, C.N. Pace, Hydrogen bonding stabilises globular proteins, Biophys. J. 71 (1996) 2033–2039.

[19] G.D. Rose, R. Wolfenden, Hydrogen-bonding, hydrophobicity, packing and protein folding, Annu. Rev. Biophys. Biomol. Struct. 22 (1993) 381–415.

[20] E.S. Domalski, E.D. Hearing, Heat capacities and entropies of organic compounds in the condensed phase. Volume III, J. Phys. Chem. Ref. Data 25 (1996) 1–525.

[21] J.S. Chickos, D.G. Hesse, J.F. Liebman, A group additivity approach for the estimation of heat capacities of organic liquids and solids at 298 K, Struct. Chem. 4 (1993) 261–269.

[22] W. Pfeil, Protein Stability and Folding: A Collection of Thermodynamic Data, Springer-Verlag, Berlin and Heidelberg, 1998.

[23] F.M. Richards, The interpretation of protein structures: total volume, group volume distributions and packing density, J. Mol. Biol. 82 (1974) 1–14.

[24] A. Cooper, Thermodynamic fluctuations in protein molecules, Proc. Natl. Acad. Sci. USA 73 (1976) 2740–2741.

[25] J. Tsai, R. Taylor, C. Chothia, M. Gerstein, The packing density in proteins: standard radii and volumes, J. Mol. Biol. 290 (1999) 253–266.

[26] Y. Zhou, D. Vitkup, M. Karplus, Native proteins are surface molten solids: application of the Lindemann criterion for solid versus liquid state, J. Mol. Biol. 285 (1999) 1371–1375.

[27] I.K. McDonald, J.M. Thornton, Satisfying hydrogen bonding potential in proteins, J. Mol. Biol. 238 (1994) 777–793.

[28] D.F. Stickle, L.G. Presta, K.A. Dill, G.D. Rose, Hydrogen bonding in globular proteins, J. Mol. Biol. 226 (1992) 1143–1159.

[29] M. Petukhov, D. Cregut, C.M. Soares, L. Serrano, Local water bridges and protein conformational stability, Protein Sci. 8 (1999) 1982–1989.

[30] F.H. Stillinger, Eccentricities of an everyday substance (book review), Nature 401 (1999) 850–851.

[31] I.M. Klotz, S.B. Farnham, Stability of an amide–hydrogen bond in an apolar environment, Biochemistry 7 (1968) 3879–3882.

[32] G.C. Kresheck, I.M. Klotz, The thermodynamics of transfer of amides from an apolar to an aqueous solution, Biochemistry 8 (1969) 8–12.

[33] K.P. Murphy, E. Freire, Thermodynamics of structural stability and cooperative folding behavior in proteins, Adv. Protein Chem. 43 (1992) 313–361.

[34] H.S. Chan, S. Bromberg, K.A. Dill, Models of cooperativity in protein folding, Phil. Trans. R. Soc. Lond. B 348 (1995) 61–70.

[35] M.J. Parker, S. Marqusee, The cooperativity of burst phase reactions explored, J. Mol. Biol. 293 (1999) 1195–1210.

[36] P.L. Privalov, Cold denaturation of proteins, Crit. Rev. Biochem. Mol. Biol. 25 (1990) 281–305.

[37] F. Franks, Protein destabilization at low temperatures, Adv. Protein Chem. 46 (1995) 105–139.

[38] S.M. Habermann, K.P. Murphy, Energetics of hydrogen bonding in proteins: a model compound study, Protein Sci. 5 (1996) 1229–1239.

[39] D. Eisenberg, W. Kauzmann, The Structure and Properties of Water, Oxford University Press, 1969.

[40] G.I. Makhatadze, P.L. Privalov, Energetics of protein structure, Adv. Protein Chem. 47 (1995) 307–425.

[41] G.I. Makhatadze, Heat capacities of amino acids, peptides and proteins, Biophys. Chem. 71 (1998) 133–156.

[42] R.L. Baldwin, Temperature dependence of the hydrophobic interaction in protein folding, Proc. Natl. Acad. Sci. USA 83 (1986) 8069–8072.

[43] B. Lee, Isoenthalpic and isoentropic temperatures and the thermodynamics of protein denaturation, Proc. Natl. Acad. Sci. USA 88 (1991) 5154–5158.

[44] K.P. Murphy, P.L. Privalov, S.J. Gill, Common features of protein unfolding and dissolution of hydrophobic compounds, Science 247 (1990) 559–561.

[45] D. Xu, S.L. Lin, R. Nussinov, Protein binding versus protein folding: the role of hydrophilic bridges in protein associations, J. Mol. Biol. 265 (1997) 68–84.

[46] T. Lundbäck, H. Hansson, S. Knapp, R. Ladenstein, T. Härd, Thermodynamic characterization of non-sequence-specific DNA-binding by the Sso7d protein from *Sulfolobus solfataricus*, J. Mol. Biol. 276 (1998) 775–786.

[47] A.J. Maynard, G.J. Sharman, M.S. Searle, Origin of β-hairpin stability in solution: structural and thermodynamic analysis of the folding of a model peptide supports hydrophobic stabilization in water, J. Am. Chem. Soc. 120 (1998) 1996–2007.

[48] J.W. Taylor, N.J. Greenfield, B. Wu, P.L. Privalov, A

calorimetric study of the folding–unfolding of an α-helix with covalently closed N and C-terminal loops, J. Mol. Biol. 291 (1999) 965–976.
[49] R. Komsa-Penkova, R. Koynova, G. Kostov, B.G. Tenchov, Thermal stability of calf skin collagen type I in salt solutions, Biochim. Biophys. Acta 1297 (1996) 171–181.
[50] S. Kolde, X. Huang, K. Link, A. Koide, Z. Bu, D.M. Engelman, Design of single-layer β-sheets without a hydrophobic core, Nature 403 (2000) 456–460.
[51] L. Banaszak, N. Winter, Z. Lu, D.A. Bernlohr, S. Cowan, T.A. Jones, Lipid-binding proteins: a family of fatty acid and retinoid transport proteins, Adv. Protein Chem. 45 (1994) 89–151.
[52] E.W. Washburn (Ed.), International Critical Tables of Numerical Data, Physics, Chemistry and Technology, vol. 5 (1929), Published for the National Research Council by the McGraw-Hill Book Company, New York, London, 1926–1933.