Hydrogen-bonding classes in proteins and their contribution to the unfolding reaction

RAFFAELE RAGONE

Dipartimento di Biochimica e Biofisica and CRISCEB, Seconda Università di Napoli, 80138 Naples, Italy

(RECEIVED MARCH 9, 2001; FINAL REVISION JULY 10, 2001; ACCEPTED JULY 12, 2001)

Abstract

This paper proposes to assess hydrogen-bonding contributions to the protein stability, using a set of model proteins for which both X-ray structures and calorimetric unfolding data are known. Pertinent thermodynamic quantities are first estimated according to a recent model of protein energetics based on the dissolution of alkyl amides. Then it is shown that the overall free energy of hydrogen-bond formation accounts for a hydrogen-bonding propensity close to helix-forming tendencies previously found for individual amino acids. This allows us to simulate the melting curve of an alanine-rich helical 50-mer with good precision. Thereafter, hydrogen-bonding enthalpies and entropies are expressed as linear combinations of backbone-backbone, backbone-side-chain, side-chain-backbone, and side-chain-side-chain donor-acceptor contributions. On this basis, each of the four components shows a different free energy versus temperature trend. It appears that structural preference for side-chain-side-chain hydrogen bonding plays a major role in stabilizing proteins at elevated temperatures.

Keywords: Group additivity; hydrophobic interaction; *N*-alkyl amides; packing interactions; protein thermodynamics; thermostability

Studies on the role played by hydrogen bonding in stabilizing protein structures are in some way linked to the more general problem of protein thermostability. In this field, two main categories have been growing so far: those analyzing the occurrence and the location in protein structures of a variety of interactions, including hydrogen bonds, and those examining the thermodynamics of model substances.

The first category is made up of comparative studies on specific mesophilic and thermophilic homologs as well as systematic analyses on enlarged, although limited, protein sets (Stickle et al. 1992; Querol et al. 1996; Petukhov et al. 1997; Vogt and Argos 1997; Vogt et al. 1997; Facchiano et al. 1998; Kumar and Nussinov 1999; Kumar et al. 2000; Szilágyi and Závodszy 2000). In any case, the limited amount of data presently available does not allow a general conclusion to be drawn. On the other hand, there are several

theoretical efforts devoted to quantify hydrogen-bonding thermodynamics. In these studies, disagreement among researchers is the rule more than the exception. In fact, a familiar argument is that the net effect produced by hydrogen bonds approximates zero, on consideration that groups involved in hydrogen bonding in the folded state can establish the same kind of interaction with water molecules in the unfolded state. There are, however, researchers who favor a role stabilizing the native structure, although to an extent that depends on thermodynamic data used as reference (Schellman 1955; Fersht et al. 1985; Murphy and Gill 1991; Serrano et al. 1992; Shirley et al. 1992; Makhataдзе and Privalov 1993; Habermann and Murphy 1996; Myers and Pace 1996; Takano et al. 1999), or even investigators who estimate that hydrogen bonding can make an unfavorable contribution (Dill 1990; Yang and Honig 1995a,b).

Some theoretical models of protein unfolding consider thermodynamic properties of solid oligopeptides (Graziano et al. 1996) or diketopyperazines (Habermann and Murphy 1996). These studies are in substantial agreement with earlier calorimetric studies on a number of proteins (Murphy et al. 1990) and on the α -helix to coil transition of an alanine-

Reprint request to: Raffaele Ragone, Dipartimento di Biochimica e Biofisica and CRISCEB, Seconda Università di Napoli, Via Costantinopoli 16, 80138 Naples, Italy; e-mail: ragone@unina2.it; fax: 390-81-294136.

Article and publication are at <http://www.proteinscience.org/cgi/doi/10.1101/ps.09201>.

rich 50-mer (Scholtz et al. 1991), suggesting that hydrogen bonding stabilizes native protein structures enthalpically. Along this track, the present study proposes to estimate the contribution made by four different subtypes of hydrogen bonds to protein stability, using a set of proteins for which both three-dimensional structures and thermodynamic data are available. On the hypothesis that the identification of unfolding quantities performed by the thermodynamics of alkyl amide dissolution is correct, the major outcome is that the nature of the donor-acceptor pair differently weighs with stabilization of the native over the thermally denatured state. This could be useful in understanding the thermodynamic basis of thermal stability.

Theoretical premise

The scheme adopted to characterize the hydrogen-bonding thermodynamics is based on the dissolution energetics of a set of liquid *N*-alkyl amides (Konicek and Wadsö 1971; Sköld et al. 1976), partly following the theoretical formalism developed by Lee (1991). The use of these compounds was previously suggested by Record and coworkers (Spolar et al. 1992), on consideration that the dissolution of liquid hydrocarbons, whenever satisfactorily mimicking the solvation of nonpolar surfaces buried in the protein interior, cannot take into account the interaction of the peptide backbone with water. Subsequently, the liquid amide dissolution was used to justify some regularities occurring in the thermal unfolding of a number of globular proteins (Ragone and Colonna 1994a,b, 1995; Ragone et al. 1995, 1996). The major outcomes of this analysis are listed below.

First, it was shown that the solvation enthalpy of nonpolar moieties becomes zero at ~ 309 K. This is not far from room temperature, which is the temperature typical of the liquid hydrocarbon transfer (Gill and Wadsö 1976; Sturtevant 1977; Baldwin 1986; Muller 1993; Schellman 1997). Second, it was proposed that the enthalpy associated with the transfer of the peptide group into water (~ -13 kJ/mole; Ragone et al. 1996) shifts the temperature at which the total solvation enthalpy is zero from $T_{np} \approx 309$ K to $T_h^* \approx 376$ K (Ragone and Colonna 1994b; Ragone et al. 1995). The close vicinity between this last temperature and the so-called enthalpy convergence temperature observed for proteins (Privalov 1979; Privalov and Gill 1988; Murphy et al. 1990) led us to conclude that polar and nonpolar solvation enthalpies of unfolding exactly counterbalance each other at $T_h^* \approx 376$ K. Third, it was found that if the enthalpy for the overall solvation process vanishes at $T_h^* \approx 376$ K, the entropy for the same process must become zero as well but at $T_s^* \approx 385$ K (Ragone and Colonna 1994a,b, 1995; Ragone et al. 1995, 1996), that is, close to the so-called entropy convergence temperature of unfolding (Privalov 1979; Privalov and Gill 1988; Murphy et al. 1990). This result was obtained using two other temperatures typical of proteins, at which the

unfolding enthalpy and entropy intersect zero (T_h and T_s , respectively; Becktel and Schellman 1987; Muller 1993; Schellman 1997), but other investigators independently attained a similar conclusion (Baldwin and Muller 1992; Doig and Williams 1992). Therefore, in the liquid amide model T_s^* represents the temperature at which solvation entropies of polar and nonpolar moieties exactly balance with one another. It is practically coincident (R. Ragone, unpubl.) with the temperature at which the solvation entropy of liquid hydrocarbons was previously shown to vanish (Gill and Wadsö 1976; Sturtevant 1977; Baldwin 1986), given that the polar interaction is mainly enthalpic (Lee 1991).

In further support of the amide model, it is perhaps worth mentioning that ensemble average heat capacity profiles characterized by the two endothermic peaks associated with cold and heat denaturation of proteins (Graziano et al. 1997) may be generated by inserting the negative enthalpy originated by the peptide group into Ikegami's mean-field statistical thermodynamic model of protein energetics (Ikegami 1977; Kanehisa and Ikegami 1977). Accordingly, at very low temperatures, the interaction between water and the peptide moiety starts dominating the unfolding energetics and likely causes cold denaturation (Graziano et al. 1997), overshadowing the role played by the increased solubility of nonpolar groups (Creighton 1991; Dill and Stigter 1995).

On this basis, it is apparent that the amide model provides a rationale for separating temperature-dependent and residual enthalpies and entropies. The former can be attributed to water solvation, because they depend on the unfolding heat capacity change, which in turn reflects the immersion of nonpolar moieties into water. They intersect zero at T_h^* and T_s^* , respectively. At these temperatures, the residual enthalpy and entropy of unfolding (ΔH_{376} and ΔS_{385} , respectively) are thus assumed to reflect mainly the solid-like packing of proteins (Ragone and Colonna 1995; Ragone et al. 1996). ΔS_{385} is explained in terms of packing or configurational entropy effects, which agrees with other investigators (Murphy et al. 1990; Lee 1991; Murphy and Gill 1991; Muller 1993). On the other hand, the amide model postulates that ΔH_{376} is the enthalpic counterpart of ΔS_{385} . The proportionality between free energy of deformation and solvation-independent (residual) part of the unfolding free energy supports this view. In fact, both quantities are characterized with large enthalpy-entropy compensation and tend to zero at roughly the same temperature, $T_m = \Delta H_{376} / \Delta S_{385} = 353 \pm 20$ K (Morozov and Morozova 1993). Other approaches, instead, ascribe ΔH_{376} to polar interactions.

Among shortcomings that may affect the dissection procedure described above, it must be mentioned that there are different views on the existence of T_h^* and T_s^* , as discussed elsewhere (Costas et al. 1994; Vailaya and Horváth 1996; Mancera et al. 1998). Moreover, a crude estimate indicates that prediction of the unfolding thermodynamics through

additive schemes, such as that adopted by us, requires modeling of contact, transfer, or binding interactions to better than ~ 400 J/mole for molecules about the size of amino acids (Dill 1997). Finally, the role played by the various kinds of forces depends somewhat on how the unfolding free energy is broken down into individual components, as well as on the choice of compounds for modeling the unfolding phenomenon (Honig and Yang 1995).

Results

In the preceding section, it has been shown that there is both theoretical (Lee's model) and experimental (amide model) evidence that the residual ΔH_{376} and ΔS_{385} do not include solvation terms. As a consequence, these quantities are usually interpreted as an evidence that proteins are as densely packed as organic crystals and are therefore believed to arise from hydrogen bonding as well as van der Waals interactions among nonpolar hydrogens. However, it should be considered that regardless of the starting phase, a large heat capacity change invariably accompanies the solvation process of nonpolar hydrogens (to which any hydrophobic effect is usually attributed), with the consequence that their unpacking and solvation are intermingled in the temperature-dependent part of the unfolding reaction. In other words, the transfer of nonpolar hydrogens from the solid-like protein interior into water incorporates van der Waals interactions absent when a nonpolar liquid enters water (Muller 1993), thus mixing with temperature-dependent quantities. A previous analysis based on the thermodynamics of alkane melting found that "close packing of the protein interior makes only a small free energy contribution to folding, because the enthalpic gain resulting from increased dispersion interactions (relative to the liquid) is countered by the freezing of side-chain motion." It was then concluded that dispersion interactions roughly amount to $\sim 10\%$ of hydrophobicity, but they "make no net contribution to the hydrophobic effect for liquid hydrocarbons" (Nicholls et al. 1991). On the other hand, the immersion of hydrogen bonds in water is not hidden by any heat capacity change. Thus, it seems quite reasonable to assume that any residual quantity (i.e., ΔH_{376} and ΔS_{385}) can be solely ascribed to their disruption.

In this regard, estimates of thermodynamic quantities associated with solid-like packing ($\Delta G_{\text{fus}} = \Delta H_{\text{fus}} - T\Delta S_{\text{fus}}$) can be still obtained by the amide model. To this aim, enthalpic data on the dissolution of two pairs of isomeric *N*-alkyl amides with solid or liquid aggregation state are shown in Figure 1. According to the considerations reported above, packing differences among nonpolar hydrogens are ascribable to the negligible differences in the heat capacity change, which result in lines that are nearly parallel for solid and liquid amides. Then, assuming that ΔH_{fus} chiefly reflects contributions from the donor-acceptor pair present in

Fig. 1. Solution enthalpy of isomeric solid and liquid *N*-alkyl amides. The packing contribution to the dissolution process can be easily estimated, assuming that the enthalpy of the solid isomers can be dissected according to the scheme solid $\rightarrow$ liquid $\rightarrow$ water, where the liquid-to-water step is given by the liquid isomers. The fusion enthalpy, indicated by the arrows, is given by the difference between solution enthalpies of *N*-*t*-butyl-acetamide (solid circles) or 2,2,*N*-trimethyl-propanamide (solid squares) and their liquid isomers, *N*-butyl-acetamide (open circles) or *N*-methyl-pentanamide (open squares), 8.81 and 12.12 kJ/mole at 25°C, respectively, and is obtained using the original thermochemical data (Konicek and Wadsö 1971; Sköld et al. 1976).

each amide, each hydrogen bond contributes ~ 4.4 or 6.1 kJ/mole, depending on the pair of isomers. Unfortunately, the lack of solubility data does not allow us to obtain an estimate for ΔS_{fus} . However, we can find reasonable limits, considering that at room temperature ($T = 298$ K) ΔG_{fus} must be positive and that the melting temperature for *N*-*t*-butyl-acetamide or 2,2,*N*-trimethyl-propanamide should not be far from 373 K. The analysis of amide enthalpy data leads therefore to conclude that does $\Delta H_{\text{fus}}/373 \leq \Delta S_{\text{fus}} < \Delta H_{\text{fus}}/298$, that is, ΔS_{fus} is expected to span 12 to 15 or 16 to 20 J/mole per K per hydrogen bond, depending on the pair of isomers.

The above evidence favors the hypothesis that ΔH_{376} and ΔS_{385} reflect only the disruption of hydrogen bonds (Table 1), to which configurational entropy effects may be conveniently attributed. A linear regression analysis considering the total number of hydrogen bonds (N_{HB}) gives 5.26 ± 0.33 kJ/mole ($r = 0.848$) and 16.5 ± 1.0 J/mole per K ($r = 0.854$) for the enthalpy and the entropy of disrupting an average hydrogen bond, respectively, which are intriguingly close to the values estimated by alkyl amides. Assuming that $\Delta G_{\text{HB}}^{\circ} = 5.26 - T \times 0.0165$ (in units of kJ/mole) represents the free energy for disrupting a single hydrogen bond, the equilibrium constant for hydrogen bond formation equals 1.15 at 298 K, which is close to sequence independent helix propensity values (O'Neil and DeGrado 1990; Wojcik et al. 1990; Chen et al. 1992). $\Delta G_{\text{HB}}^{\circ}$ was then used to draw the unfolding curve of a hypothetical helical 50-mer (Fig. 2), which is to be compared with the

Table 1. Global census of hydrogen bonds

Protein	<i>N</i>	<i>N</i> _{NO}	<i>N</i> _{RO}	<i>N</i> _{NR}	<i>N</i> _{RR}	<i>N</i> _{HB}	<i>N</i> _{rHB}	ΔH_{376}	ΔS_{385}
Trypsin inhibitor	58	27	4	6	2	39	28	363.7	1.1
Amylase inhibitor	74	35	10	9	8	62	46	336.2	1.0
Ubiquitin	76	52	9	11	4	76	50	471.4	1.4
Cytochrome c	103	75	18	10	3	106	62	631.7	2.0
Lys25-ribonuclease T1	104	65	15	12	12	104	61	750.3	2.4
Ribonuclease A	124	74	25	13	18	130	82	775.5	2.4
Lysozyme, human	130	105	20	16	18	159	95	729.2	2.3
Staphylococcus nuclease	136	101	13	19	18	151	92	807.4	2.6
Myoglobin	153	194	17	13	20	244	129	853.9	2.7
Lysozyme, T4	164	179	17	20	26	242	126	1062.9	3.4
Papain	212	132	31	27	31	221	135	1290.4	3.9
β -Trypsin	223	123	25	26	16	190	136	1225.3	3.8
α -Chymotrypsin A	241	126	26	25	14	191	136	1253.1	4.0

Adapted from Stickle et al. (1992). Columns 1 and 2 identify each protein and its length (*N*). Columns 3 through 8 list the number of hydrogen bonds: backbone donor to backbone acceptor (*N*_{NO}), side-chain donor to backbone acceptor (*N*_{RO}), backbone donor to side-chain acceptor (*N*_{NR}), side-chain donor to side-chain acceptor (*N*_{RR}), total (*N*_{HB}), and reduced (*N*_{rHB}). Columns 9 and 10 include the assumed packing enthalpy (in units of kJ/mole) and entropy (in units of kJ/mole per K) of unfolding, evaluated according to equations 1 and 2 (see Materials and Methods), respectively.

experimental α -helix to coil transition of a 50-residue alanine-rich peptide (Scholtz et al. 1991). Because of the satisfactory agreement between theoretical and experimental curve, average free energy terms allow us to model the hydrogen-bonding energetics in a protein-like system, whenever made up of a single class of hydrogen bonds (backbone donor to backbone acceptor, in the present case),

Fig. 2. Unfolding curve of a helical 50-mer. The experimental helical fraction (f_{HB}) of the alanine-rich peptide (solid line) was obtained using a van't Hoff enthalpy (ΔH_{vH}) of 46 kJ/mole (corresponding to 11 kcal/mole) and a midpoint temperature of 314 K (Scholtz et al. 1991). The theoretical unfolding curve (dashed line) was drawn with the assumption that $\Delta G_{HB}^{\circ} = 5.26 - T \times 0.0165$ (in units of kJ/mole) represents the free energy for disrupting a single hydrogen bond. ΔG_{HB}° was then multiplied by 50, to account for length, and divided by 6, to account for deviation from a two-state reaction, according to the calorimetric to van't Hoff enthalpy ratio originally found. The midpoint temperature is ~ 319 K. Long dashed curves refer to myoglobin and ubiquitin (*mbn* and *ubq*, respectively) and were obtained using values of ΔH_{376} and ΔS_{385} reported in Table 1.

in absence of nonpolar solvation contributions. In fact, the thermal unfolding of this polypeptide occurs without any appreciable heat capacity change. However, more detailed information on the temperature dependence of hydrogen bonding in protein systems may be obtained only analyzing a few models. As an example, Figure 2 also shows how the fraction of hydrogen bonds depends on temperature for myoglobin and ubiquitin. It can be appreciated that the midpoint temperature is located around 316 and 337 K for myoglobin and ubiquitin, respectively. This difference, which may be suggested to reflect packing differences between the two proteins (Facchiano et al. 1999), cannot be explained using average free energy parameters. In the following, the different behavior of these proteins will be rationalized in terms of the different classes that contribute to the hydrogen-bonding thermodynamics.

The above analysis does not include any explicit information on the existence of hydrogen-bond networks. In fact, if the reduced number of hydrogen bonds (N_{rHB}) were to be used as the independent variable, enthalpic and entropic contributions to the protein stability would equal 8.78 ± 0.35 kJ/mole ($r = 0.934$) and 27.5 ± 1.0 J/mole per K ($r = 0.936$) per reduced hydrogen bond, respectively. As, in general, there are clusters in which each donor (acceptor) is bonded to multiple acceptors (donors), N_{rHB} represents the maximum number of hydrogen bonds possible for a given cluster if every donor-acceptor pair were constrained to be 1:1. Thus, the differences between N_{HB} and N_{rHB} -dependent figures are proportional to the co-operative enhancement caused by the existence of hydrogen bond networks (Stickle et al. 1992). It seems reasonable that this presumably reflects structural preference for specific hydrogen-bonding classes, which are likely to weigh differently

with values of ΔH_{376} and ΔS_{385} , depending on the preference induced by the three-dimensional protein arrangement. On this basis, we can make a step forward considering backbone donor-to-backbone acceptor (N_{NO}), side-chain donor-to-backbone acceptor (N_{RO}), backbone donor-to-side-chain acceptor (N_{NR}), and side-chain donor-to-side-chain acceptor (N_{RR}) hydrogen bonds (see Table 1).

Multilinear regression analyses performed by equations 3 and 4 (see Materials and Methods) suggest that a single NO-, RO-, NR-, or RR-type hydrogen bond contributes 1.859, 12.260, 29.484, or -4.199 kJ/mole to ΔH_{376} , and 6.5, 34.8, 93.4, or -15.2 J/mole per K to ΔS_{385} , respectively. It is worth reminding here that this set of parameters includes correlation (nonadditivity) among the various hydrogen-bonding contributions. In other words, it can be used if, and only if, we are to analyze systems in which the four different hydrogen-bonding subtypes are represented together, according to the identification procedure adopted by Stickley and coworkers (1992). Any other analysis, based on the assumption that these parameters are valid independently of each other, needs caution and may lead to wrong conclusions. Free energy versus temperature trends shown in Figure 3 highlight contributions of these different components to the unfolding energetics. Namely, at room temperature all hydrogen bonds but the backbone-to-backbone ones stabilize the native structure, whereas at higher temperatures all hydrogen bonds become destabilizing but the side-chain-to-side-chain ones. The relative trends displayed by the various classes should not be affected by the choice of dif-

ferent values for the residual enthalpy and entropy. On this basis, we can explain why in Figure 2 the curve relative to ubiquitin is shifted to higher temperatures by ~ 21 degrees compared with that of myoglobin. From Table 1 it can be inferred that the weight of NO-, RO-, NR-, or RR-type hydrogen bonds is 68.4%, 11.8%, 14.5%, and 5.3% for ubiquitin and 79.5%, 7.0%, 5.3%, and 8.2% for myoglobin. Overall, this implies that in this last protein, the most heat labile interactions are represented to an extent larger than $\sim 11\%$ as compared with those present in ubiquitin. In this favorable case, it appears that the higher heat denaturation temperature of ubiquitin (~ 367 K versus ~ 350 K; for references, see Facchiano et al. 1999) can be therefore entirely attributed to hydrogen-bonding differences between the two proteins.

Discussion

To date, hydrogen-bonding effects on the unfolding energetics have been characterized using a variety of approaches, including the analysis of the unfolding thermodynamics of specific proteins, site-directed mutagenesis, simplified protein models, and theoretical calculations. The analysis herein proposed is based on three-dimensional structures of proteins and assumes that the free energy for the disruption of a single hydrogen bond (ΔG_{HB}) is the linear combination of four different contributions. To this aim, pertinent thermodynamic quantities have been discriminated through dissection of the overall unfolding process into melting of the crystal-like protein core and solvation of internal moieties, including van der Waals effects linked to the solid-like packing of nonpolar hydrogens. It has been shown that the estimate of the average hydrogen-bonding strength obtained by the amide thermodynamics is in substantial agreement with that derived by considering the actual number of hydrogen bonds in proteins. The validity of this estimate has been checked generating the melting curve of a hypothetical 50-residue peptide and verifying that it is very close to that of an alanine-rich 50-mer (Scholtz et al. 1991). The small midpoint shift (see Fig. 2) could be justified, considering that the alanine 50-mer is purely helical, which does not occur with proteins. Moreover, an explanation has been given for the fact that myoglobin and ubiquitin unfold at ~ 350 and 367 K, respectively.

Assuming that each hydrogen-bonding class is equally represented, equation 6 (see Materials and Methods) allows us to calculate that the free energy for a single hydrogen bond amounts to 0.95 kJ/mole at room temperature. This value differs by at least one order of magnitude from those obtained by other investigators at room temperature (for examples, see Schellman 1955; Fersht et al. 1985; Murphy and Gill 1991; Serrano et al. 1992; Shirley et al. 1992; Fersht and Serrano 1993; Makhataadze and Privalov 1993; Habermann and Murphy 1996; Myers and Pace 1996; Ta-

Fig. 3. Contribution of hydrogen bonds to protein unfolding. Hydrogen bonds were classified as backbone donor-to-backbone acceptor (NO), side-chain donor-to-backbone acceptor (RO), backbone donor-to-side-chain acceptor (NR), side-chain donor-to-side-chain acceptor (RR) bonds. Lines were drawn according to equation 5 (see Materials and Methods), using fitting parameters thus obtained, i.e., $\Delta h_{NO} = (1.9 \pm 0.7)$, $\Delta h_{RO} = (12.3 \pm 4.6)$, $\Delta h_{NR} = (29.5 \pm 5.7)$, and $\Delta h_{RR} = (-4.2 \pm 4.5)$ kJ/mole for the enthalpy ($r = 0.978$), and $\Delta s_{NO} = (6.5 \pm 2.2)$, $\Delta s_{RO} = (34.8 \pm 15.6)$, $\Delta s_{NR} = (93.4 \pm 19.2)$, and $\Delta s_{RR} = (-15.2 \pm 15.3)$ J/mole per K for the entropy ($r = 0.975$).

kano et al. 1999). This should not depend on how the denatured state is treated, which is operationally defined here as the population of conformations produced by thermal denaturation. Other investigators have strongly considered the possibility that hydrogen bonding makes little or no net contribution to the protein stability (Dill 1990; Honig and Yang 1995; Yang and Honig, 1995a,b). Moreover, other biological systems exist for which the disruption of one hydrogen bond requires a free energy as low as ~ 0.8 kJ/mole (Kato et al. 1995). Commenting on differences among hydrogen-bonding stabilities existing in the literature, Honig and Yang (1995) concluded that "the question is to some extent one of definitions and the method used to partition free energies into individual components."

The present estimate results from the linear combination of four different contributions, all of which are represented in proteins, seemingly suggesting that the nature of the donor/acceptor pair might control the unfolding energetics. It must be also borne in mind that this does not really provide evidence that the strength of a hydrogen bond depends on whether the donor or the acceptor is in the backbone or in the side-chain, which would be unreasonable. Values reported here are to be considered averages on hydrogen-bonding geometries and lengths weighed by structural requirements imposed by the three-dimensional folding of the protein chain. It is therefore hard to compare this result to others that may be restricted to specific types of hydrogen bonds to assume linear combinations of somewhat different hydrogen-bonding classes (for example, see Habermann and Murphy 1996).

From the above analysis, it appears that the enthalpy of hydrogen bonding is the major source of discrepancy with other studies. In fact, there is substantial agreement among researchers that configurational entropy effects can be explained in terms of packing, including hydrogen bonds (Murphy et al. 1990; Lee 1991; Murphy and Gill 1991; Muller 1993). Thus, estimating in 18 J/mole per K, the entropy required to disrupt 1 mole of hydrogen bonds (Muller 1993), which is close to the weighed estimate of 17.1 J/mole per K extractable from Table 1, and using a stabilization free energy of 2.1 kJ/mole, which is the lowest stability decrease induced by removal of 1 mole of hydrogen bonds found by Fersht and Serrano (1993), we obtain an enthalpy of ~ 7.5 kJ/mole at 298 K. However, this would result in a midpoint temperature for the disruption of 1 mole of hydrogen bonds of ~ 414 K. Even without considering contributions originating from the solvation of the protein interior, this value is well beyond the heat denaturation temperature of most monomeric proteins.

Nevertheless, it may be useful to address this discrepancy by comparing the present work with a specific previous study, in which the investigators were interested in explaining the origin of the thermal stability of proteins from thermophilic sources (Vogt et al. 1997). In their comparative

analysis of structural features of homologous thermophilic and mesophilic proteins, Argos and coworkers estimated that "for each 10°C increase in living temperature, a mean total of ~ 13 hydrogen bonds and salt links is added." In addition, "assuming a stabilization gain of 0.5 kcal/mole" (2.1 kJ/mole) "per bond formed and an average increase of 40°C in thermal stability," they obtained that " ~ 26 kcal/mole" (108.8 kJ/mole) "would separate the folded free energies of mesophiles and thermophiles" (Vogt et al. 1997). Instead, the value of 0.95 kJ/mole gives 49.4 kJ/mole. These different estimates are to be compared with the measured difference range of 20.9 to 83.7 kJ/mole (Vogt et al. 1997). It seems evident that the analysis presented here leads to rather good agreement with the experimental finding, whereas even the most favorable value of 0.5 kcal/mole adopted by Vogt and coworkers, according to Fersht and Serrano (1993), produces an overestimate.

What is more, the relative temperature dependence of the four hydrogen-bonding contributions identified should not be affected by different estimates of thermodynamic quantities. On this basis, it is apparent that the side-chain donor-to-side-chain acceptor class is the best candidate to preserve proteins from unfolding on rising temperature. It is increasingly native stabilizing at elevated temperatures. In particular, it seems reasonable to speculate that the high frequency of ion pairs (Stickle et al. 1992) plays a role in determining this peculiar temperature dependence. Although there is incidental evidence that coulombic interactions tend to favor the unfolded state (Honig and Nicholls 1995), in comparative analyses of structural parameters in representatives of families of homologous thermophilic and mesophilic proteins, it was noticed that an increase in ion pairs occurs in thermostable proteins (Szilágyi and Závodszy 2000), often paralleling an increase in hydrogen bonds (Scandurra et al. 1998, 2000; Kumar et al. 2000). This strengthens findings reported here, which are based on mesophilic proteins and therefore are not biased by any information on thermophilic systems. Another piece of evidence seems to favor the above hypothesis. The increased relative stabilities of large hyperthermophilic proteins at high temperatures can be attributed to a large difference between mean net charges on the folded and unfolded proteins (Helgeson 1999). Larger populations of charged residues have actually been found in hyperthermophilic enzymes than in mesophilic proteins of smaller size. Therefore, it has been argued that a large population of ionized residues on the outer surface of a protein is by itself sufficient to increase the stability of large proteins at high temperatures (Helgeson 1999). This is consistent with earlier suggestions (Perutz and Raidt 1975). However, this structural preference is concomitant with other adjustments of the protein architecture (Scandurra et al. 1998, 2000; Kumar et al. 2000). At present no clear-cut conclusion can be drawn, but this study agrees with recent results (Vetriani et al. 1998; Grimsley et al. 1999) in concluding

that the thermostability-driving role played by coulombic interactions appears to deserve thorough attention in the future.

Materials and methods

The global census performed on proteins listed in Table 1 was used to evaluate the contribution made by hydrogen bonds to protein stability (Stickle et al. 1992). According to the dissection procedure outlined in the Theoretical Premise section, residual enthalpies and entropies of unfolding were evaluated through the equations

$$\Delta H_{376} = \Delta H_{T^\circ} + \Delta C_p(T_h^* - T^\circ) \quad (1)$$

$$\Delta S_{385} = \Delta S_{T^\circ} + \Delta C_p(\ln T_s^* - \ln T^\circ) \quad (2)$$

using calorimetric data previously published (for references, see Ragone et al. 1996). Otherwise, calculations were performed using data from the ProTherm database, which is freely accessible at <http://www.rtc.riken.go.jp/protherm.html> (Gromiha et al. 1999). In the above expressions, $T_h^* = 376$ K and $T_s^* = 385$ K represent the temperatures at which the total (polar plus nonpolar) solvation enthalpy and entropy go to zero, respectively. ΔH_{376} and ΔS_{385} are thus assumed to represent enthalpic and entropic contributions to the packing free energy, ΔH_{T° and ΔS_{T° are the unfolding enthalpy and entropy at an arbitrary reference temperature T° , and ΔC_p is the molar heat capacity change of unfolding at constant pressure.

Although it is known that ΔC_p depends on the temperature, calculations were performed assuming its constancy, as made in the original calorimetric studies. Furthermore, there are no data available on the temperature dependence of the heat capacity for the dissolution of *N*-alkyl amides used as model compounds. In any case, in the range of temperatures experimentally accessible at atmospheric pressure, equations 1 and 2 result in estimates that do not differ appreciably from those obtained through a temperature-dependent ΔC_p . Thus, even very thorough investigators often approximate ΔC_p as being temperature independent (for example, see Becktel and Schellman 1987; Schöppe et al. 1997). It is also worth noting that a linear dependence of the heat capacity with temperature would result in the coincidence of T_h^* and T_s^* (Vailaya and Horváth 1996). In turn, this would imply that protein folding is opposed by desolvation at every temperature (for example, see Murphy et al. 1990). This is not likely to be the case, at least at $<100^\circ\text{C}$, at which separation of nonpolar phase from water is a strongly favored process (Schellman 1997). Finally, the assumption of a constant heat capacity change does not affect estimates of the relative stabilities displayed by different hydrogen-bonding classes.

ΔH_{376} and ΔS_{385} were expressed as linear combinations of four classes of hydrogen bonds (see Table 1), implementing the equations

$$\Delta H_{376} = \sum \Delta h_i \times N_i \quad (3)$$

$$\Delta S_{385} = \sum \Delta s_i \times N_i \quad (4)$$

in the package Scientist for Windows, version 2.0 (MicroMath Software). In these equations, the parameters Δh_i and Δs_i represent the enthalpic and entropic contribution made by a single class *i* hydrogen bond, inclusive of any structure-dependent preference, respectively, and N_i stays for the number of these hydrogen bonds (see Table 1). Free energy estimates (Δg_i) were then obtained by

$$\Delta g_i = \Delta h_i - T\Delta s_i \quad (5)$$

Finally, the free energy (ΔG_{HB}) for disrupting one average hydrogen bond was evaluated through

$$\Delta G_{\text{HB}} = (\sum \Delta h_i - T \times \sum \Delta s_i) / 4 \quad (6)$$

where it is assumed that ΔG_{HB} is the arithmetic mean of the four hydrogen-bonding contributions identified.

Acknowledgments

This study was financed with funds of the PRIN (National Interest Research Projects) Project on Structural Determinants of Stability, Folding, and Activity of Thermophilic Enzymes from the Italian Ministry of University and Scientific and Technological Research (MURST Grant No. 9905187893_002).

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

References

- Baldwin, R.L. 1986. Temperature dependence of the hydrophobic interaction in protein folding. *Proc. Natl. Acad. Sci.* **83**: 8069–8072.
- Baldwin, R.L. and Muller, N. 1992. Relation between the convergence temperatures T_h^* and T_s^* in protein unfolding. *Proc. Natl. Acad. Sci.* **89**: 7110–7113.
- Becktel, W.J. and Schellman, J.A. 1987. Protein stability curves. *Biopolymers* **26**: 1859–1877.
- Chen, C.-C., Zhu, Y., King, J.A., and Evans, L.B. 1992. A molecular thermodynamic approach to predict the secondary structure of homopolypeptides in aqueous systems. *Biopolymers* **32**: 1375–1392.
- Costas, M., Kronberg, B., and Silveston, R. 1994. General thermodynamic analysis of the dissolution of non-polar molecules into water. *J. Chem. Soc., Faraday Trans.* **90**: 1513–1522.
- Creighton, T.E. 1991. Stability of folded conformations. *Curr. Opin. Struct. Biol.* **1**: 5–16.
- Dill, K.A. 1990. Dominant forces in protein folding. *Biochemistry* **29**: 7133–7155.
- . 1997. Additivity principles in biochemistry. *J. Biol. Chem.* **272**: 701–704.
- Dill, K.A. and Stigter, D. 1995. Modeling protein stability as heteropolymer collapse. *Adv. Protein Chem.* **46**: 59–104.
- Doig, A.J. and Williams, D.H. 1992. Why water soluble, compact, globular proteins have similar specific enthalpies of unfolding at 110°C . *Biochemistry* **31**: 9371–9375.
- Facchiano, A.M., Colonna, G., and Ragone, R. 1998. Helix stabilizing factors and stabilization of thermophilic proteins: An X-ray-based study. *Protein Eng.* **11**: 753–760.
- . 1999. An analysis of calorimetric data on mesophilic proteins and its implications for stability to elevated temperatures. *Curr. Top. Pept. Protein Res.* **3**: 101–109.
- Fersht, A.R. and Serrano, L. 1993. Principles of protein stability derived from protein engineering experiments. *Curr. Opin. Struct. Biol.* **3**: 75–83.
- Fersht, A.R., Shi, J.-P., Knill-Jones, J., Lowe, D.M., Wilkinson, A.J., Blow, D.M., Brick, P., Carter, P., Waye, M.M.Y., and Winter, G. 1985. Hydrogen bonding and biological specificity analysed by protein engineering. *Nature* **314**: 235–238.
- Gill, S.J. and Wadsö, I. 1976. An equation of state describing hydrophobic interactions. *Proc. Natl. Acad. Sci.* **73**: 2955–2958.
- Graziano, G., Catanzano, F., Del Vecchio, P., Giancola, C., and Barone, G. 1996. Thermodynamic stability of globular proteins: A reliable model from small molecule studies. *Gazz. Chim. Ital.* **126**: 559–567.
- Graziano, G., Catanzano, F., Riccio, A., and Barone, G. 1997. A reassessment of the molecular origin of cold denaturation. *J. Biochem.* **122**: 395–401.
- Grimsley, G.R., Shaw, K.L., Fee, L.R., Alston, R.W., Huyghues-Despointes, B.M.P., Thurkill, R.L., Scholtz, J.M., and Pace, C.N. 1999. Increasing pro-

- tein stability by altering long-range coulombic interactions. *Protein Sci.* **8**: 1843–1849.
- Gromiha, M.M., An, J., Kono, H., Oobatake, M., Uedaira, H., and Sarai, A. 1999. ProTherm: Thermodynamic database for proteins and mutants. *Nucleic Acids Res.* **27**: 286–288.
- Habermann, S.M. and Murphy, K.P. 1996. Energetics of hydrogen bonding in proteins: A model compound study. *Protein Sci.* **5**: 1229–1239.
- Helgeson, H.C. 1999. Thermodynamic prediction of the relative stabilities of hyperthermophilic enzymes. In *Chemical thermodynamics* (ed. T. M. Lechter), pp. 301–312. Blackwell Science, Oxford.
- Honig, B. and Nicholls, A. 1995. Classical electrostatics in biology and chemistry. *Science* **268**: 1144–1149.
- Honig, B. and Yang, A.-S. 1995. Free energy balance in protein folding. *Adv. Protein Chem.* **46**: 27–58.
- Ikegami, A. 1977. Structural changes and fluctuations of proteins: I. A statistical thermodynamic model. *Biophys. Chem.* **6**: 117–130.
- Kanehisa, M.I. and Ikegami, A. 1977. Structural changes and fluctuations of proteins: II. Analysis of the denaturation of globular proteins. *Biophys. Chem.* **6**: 131–149.
- Kato, Y., Conn, M.M., and Rebek, Jr., J. 1995. Hydrogen bonding in water using synthetic receptors. *Proc. Natl. Acad. Sci.* **92**: 1208–1212.
- Konicek, J. and Wadsö, I. 1971. Thermochemical properties of some carboxylic acids, amines and *N*-substituted amides in aqueous solution. *Acta Chem. Scand.* **25**: 1541–1551.
- Kumar, S. and Nussinov, R. 1999. Salt bridge stability in monomeric proteins. *J. Mol. Biol.* **293**: 1241–1255.
- Kumar, S., Tsai, C.-J., and Nussinov, R. 2000. Factors enhancing protein thermostability. *Protein Eng.* **13**: 179–191.
- Lee, B. 1991. Isoenthalpic and isentropic temperatures and the thermodynamics of protein denaturation. *Proc. Natl. Acad. Sci.* **88**: 5154–5158.
- Makhatadze, G.I. and Privalov, P.L. 1993. Contribution of hydration to protein folding thermodynamics: I. The enthalpy of hydration. *J. Mol. Biol.* **232**: 639–659.
- Mancera, R.L., Kronberg, B., Costas, M., and Silveston, R. 1998. Thermodynamics of the hydration of non-polar substances. *Biophys. Chem.* **70**: 57–63.
- Morozov, V.N. and Morozova, T.Y. 1993. Elasticity of globular proteins: The relation between mechanics, thermodynamics and mobility. *J. Biomol. Struct. Dyn.* **11**: 459–481.
- Muller, N. 1993. Hydrophobicity and stability for a family of model proteins. *Biopolymers* **33**: 1185–1193.
- Murphy, K.P., Privalov, P.L., and Gill, S.J. 1990. Common features of protein unfolding and dissolution of hydrophobic compounds. *Science* **247**: 559–561.
- Murphy, K.P. and Gill, S.J. 1991. Solid model compounds and the thermodynamics of protein unfolding. *J. Mol. Biol.* **222**: 699–709.
- Myers, J.K. and Pace, C.N. 1996. Hydrogen bonding stabilizes globular proteins. *Biophys. J.* **71**: 2033–2039.
- Nicholls, A., Sharp, K.A., and Honig, B. 1991. Protein folding and association: Insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11**: 281–296.
- O'Neil, K.T. and DeGrado, W.F. 1990. A thermodynamic scale for the helix-forming tendencies of the commonly occurring amino acids. *Science* **250**: 646–651.
- Perutz, M.F. and Raidt, H. 1975. Stereochemical basis of heat stability in bacterial ferredoxins and in hemoglobin A2. *Nature* **255**: 256–259.
- Petukhov, M., Kil, Y., Kuramitsu, S., and Lanzov, V. 1997. Insights into thermal resistance of proteins from the intrinsic stability of their α -helices. *Proteins* **29**: 309–320.
- Privalov, P.L. 1979. Stability of small globular proteins. *Adv. Protein Chem.* **33**: 167–237.
- Privalov, P.L. and Gill, S.J. 1988. Stability of protein structure and hydrophobic interaction. *Adv. Protein Chem.* **39**: 191–234.
- Querol, E., Perez-Pons, J.A., and Mozo-Villarias, A. 1996. Analysis of protein conformational characteristics related to thermostability. *Protein Eng.* **9**: 265–271.
- Ragone, R. and Colonna, G. 1994a. Enthalpy-entropy balance and convergence temperatures in protein unfolding. *J. Biol. Chem.* **269**: 4047–4049.
- . 1994b. The role of conditional hydration on the thermodynamics of protein folding. *J. Am. Chem. Soc.* **116**: 2677–2678.
- . 1995. Do globular proteins require some structural peculiarity to best function at high temperatures? *J. Am. Chem. Soc.* **117**: 16–20.
- Ragone, R., Stiuso, P., and Colonna, G. 1995. Enthalpy convergence temperatures: Proteins and model compounds. *Thermochim. Acta* **251**: 371–377.
- . 1996. Water transfer energetics and solid-like packing of globular proteins. *Proteins* **24**: 388–393.
- Scandurra, R., Consalvi, V., Chiaraluce, R., Politi, L., and Engel, P.C. 1998. Protein thermostability in extremophiles. *Biochimie* **80**: 933–941.
- . 2000. Protein stability in extremophilic archaea. *Front. Biosci.* **5**: d787–795.
- Schellman, J.A. 1955. Thermodynamics of urea solutions and the heat of formation of the peptide hydrogen bond. *C. R. Trav. Lab. Carlsberg, Ser. Chim.* **29**: 223–229.
- . 1997. Temperature, stability, and the hydrophobic interaction. *Biophys. J.* **73**: 2960–2964.
- Scholtz, J.M., Marqusee, S., Baldwin, R.L., York, E.J., Stewart, J.M., Santoro, M., and Bolen, D.W. 1991. Calorimetric determination of the enthalpy change for the α -helix to coil transition of an alanine peptide in water. *Proc. Natl. Acad. Sci.* **88**: 2854–2858.
- Schöppe, A., Hinz, H.-J., Agashe, V.R., Ramachandran, S., and Udgaonkar, J.B. 1997. DSC studies of the conformational stability of barstar wild-type. *Protein Sci.* **6**: 2196–2202.
- Serrano, L., Kellis, J.T., Cann, P., Matouschek, A., and Fersht, A. 1992. The folding of an enzyme: II. Structure of barnase and the contribution of different interactions to protein stability. *J. Mol. Biol.* **224**: 783–804.
- Shirley, B.A., Stanssens, P., Hahn, U., and Pace, C.N. 1992. Contribution of hydrogen bonding to the conformational stability of ribonuclease T1. *Biochemistry* **31**: 725–732.
- Sköld, R., Suurkuusk, J., and Wadsö, I. 1976. Thermochemistry of solutions of hydrocarbon model compounds: 7. Aqueous solutions of some amides, *t*-butanol and pentanol. *J. Chem. Thermodyn.* **8**: 1075–1080.
- Spolar, R.S., Livingstone, J.R., and Record, Jr., M.T. 1992. 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**: 3947–3955.
- Stickle, D.F., Presta, L.G., Dill, K.A. and Rose, G.D. 1992. Hydrogen bonding in globular proteins. *J. Mol. Biol.* **226**: 1143–1159.
- Sturtevant, J.M. 1977. Heat capacity and entropy changes in processes involving proteins. *Proc. Natl. Acad. Sci.* **74**: 2236–2240.
- Szilágyi, A. and Závodszky, P. 2000. Structural differences between mesophilic, moderately thermophilic and extremely thermophilic protein subunits: Results of a comprehensive survey. *Structure* **8**: 493–504.
- Takano, K., Yamagata, Y., Kubota, M., Funahashi, J., Fujii, S., and Yutani, K. 1999. Contribution of hydrogen bonds to the conformational stability of human lysozyme: Calorimetry and X-ray analysis of six Ser→Ala mutants. *Biochemistry* **38**: 6623–6629.
- Vailaya, A. and Horváth, C. 1996. Enthalpy-entropy compensation in hydrophobic interaction chromatography. *J. Phys. Chem.* **100**: 2447–2455.
- Vetriani, C., Maeder, D.L., Tolliday, N., Yip, K.S.-P., Stillman, T.J., Britton, K.L., Rice, D.W., Klump, H.H., and Robb, F.T. 1998. Protein thermostability above 100°C: A key role for ionic interactions. *Proc. Natl. Acad. Sci.* **95**: 12300–12305.
- Vogt, G. and Argos, P. 1997. Protein thermal stability: Hydrogen bonds or internal packing? *Fold. Des.* **2**: S40–46.
- Vogt, G., Woell, S., and Argos, P. 1997. Protein thermal stability, hydrogen bonds, and ion pairs. *J. Mol. Biol.* **269**: 631–643.
- Yang, A.S. and Honig, B. 1995a. Free energy determinants of secondary structure formation: I. α -Helices. *J. Mol. Biol.* **252**: 351–365.
- . 1995b. Free energy determinants of secondary structure formation: II. Antiparallel β -sheets. *J. Mol. Biol.* **252**: 366–376.
- Wojcik, J., Altmann, K.H., and Scheraga, H.A. 1990. Helix-coil stability constants for the naturally occurring amino acids in water. XXIV. *Biopolymers* **30**: 121–134.