Temperature Range of Thermodynamic Stability for the Native State of Reversible Two-State Proteins†

Sandeep Kumar,‡ Chung Jung Tsai,§ and Ruth Nussinov*,§,||

Laboratory of Experimental and Computational Biology, National Cancer Institute-Frederick, Building 469, Room 151, Frederick, Maryland 21702, Basic Research Program, SAIC-Frederick, Inc., and Laboratory of Experimental and Computational Biology, National Cancer Institute-Frederick, Building 469, Room 151, Frederick, Maryland 21702, and Sackler Institute of Molecular Medicine, Department of Human Genetics and Molecular Medicine, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv 69978, Israel

Received November 18, 2002; Revised Manuscript Received March 11, 2003

ABSTRACT: The difference between the heat (T_G) and the cold (T_G') denaturation temperatures defines the temperature range (T_{Range}) over which the native state of a reversible two-state protein is thermodynamically stable. We have performed a correlation analysis for thermodynamic parameters in a selected data set of structurally nonhomologous single-domain reversible two-state proteins. We find that the temperature range is negatively correlated with the protein size and with the heat capacity change (ΔC_p) but is positively correlated with the maximal protein stability [$\Delta G(T_S)$]. The correlation between the temperature range and maximal protein stability becomes highly significant upon normalization of the maximal protein stability with protein size. The melting temperature (T_G) also shows a negative correlation with protein size. Consistently, T_G and T_G' show opposite correlations with ΔC_p, indicating a dependence of the T_{Range} on the curvature of the protein stability curve. Substitution of proteins in our data set with their homologues and arbitrary addition or removal of a protein in the data set do not affect the outcome of our analysis. Simulations of the thermodynamic data further indicate that T_{Range} is more sensitive to variations in curvature than to the slope of the protein stability curve. The hydrophobic effect in single domains is the principal reason for these observations. Our results imply that larger proteins may be stable over narrower temperature ranges and that smaller proteins may have higher melting temperatures, suggesting why protein structures often differentiate into multiple substructures with different hydrophobic cores. Our results have interesting implications for protein thermostability.

A reversible two-state folding ⇌ unfolding mechanism is the simplest mechanism of protein folding. The thermodynamic stability of a two-state protein varies with temperature, pH, solvent, the presence of chemical denaturants, and salt concentration. The temperature-dependent variation in the thermodynamic stability of a two-state protein can be studied using the Gibbs−Helmholtz equation (*1*, *2*):

$$\Delta G(T) = \Delta H_G(1 - T/T_G) - \Delta C_p[(T_G - T) + T\ln(T/T_G)] \quad (1)$$

where $\Delta G(T)$[1] is the Gibbs free energy change between the denatured (D) and native (N) states of the protein at a given temperature T, ΔH_G is the enthalpy change between the two states at the melting temperature (T_G), and ΔC_p is the heat capacity change between the two states. ΔH_G, ΔC_p, and T_G can be determined using calorimetric and spectroscopic experiments. A plot of $\Delta G(T)$ versus T yields a skewed parabola-shaped protein stability curve (e.g., see Figure 1).

The hydrophobic effect (*3*) is the major force driving protein folding (*4*). However, the relative estimates of the energetic contributions of hydrophobicity and electrostatics to protein stability in the native state are controversial. A back of the envelope calculation by Schellman (*5*) showed that the hydrophobic effect is responsible for ∼75% of the protein stability. On the other hand, Pace et al. (*6*−*8*) have suggested that both hydrophobic and electrostatic effects make large but comparable energetic contributions to protein stability. By performing statistical analysis of experimental

† The research of R.N. in Israel has been supported by the Center of Excellence in Geometric Computing and its Applications, funded by the Israel Science Foundation, by a Ministry of Science grant, by a Tel Aviv University Basic Research grant, and by Adams Brain Center. This project has been funded in whole or in part with federal funds from the National Cancer Institute, National Institutes of Health, under Contract NO1-CO-12400.

* To whom correspondence should be addressed: NCI-FCRF Bldg 469, Rm 151, Frederick, MD 21702. Telephone: (301) 846-5579. Fax: (301) 846-5598. E-mail: ruthn@ncifcrf.gov.

‡ Laboratory of Experimental and Computational Biology, National Cancer Institute-Frederick.

§ SAIC-Frederick, Inc., and Laboratory of Experimental and Computational Biology, National Cancer Institute-Frederick.

|| Tel Aviv University.

[1] Abbreviations: T_G, heat denaturation (melting) temperature; T_G', cold denaturation temperature; T_S, temperature of maximal protein stability; ΔH_G, molar enthalpy change between the native and denatured states of a protein at T_G; ΔS_G, molar entropy change between the native and denatured states of a protein at T_G; ΔC_p, molar heat capacity change between the native and denatured states of a protein; $\Delta G(T)$, molar Gibbs free energy change between the native and denatured states of a protein at temperature T; $\Delta G(T_S)$, molar Gibbs free energy change between the native and denatured states of a protein at T_S; T_{Range}, temperature range of thermodynamic stability for a reversible two-state protein; $-\Delta S_G$, slope of the protein stability curve at T_G; $-\Delta C_p/T_G$, curvature of the protein stability curve at T_G; r, linear correlation coefficient.

10.1021/bi027184+ CCC: $25.00
Published on Web 04/11/2003

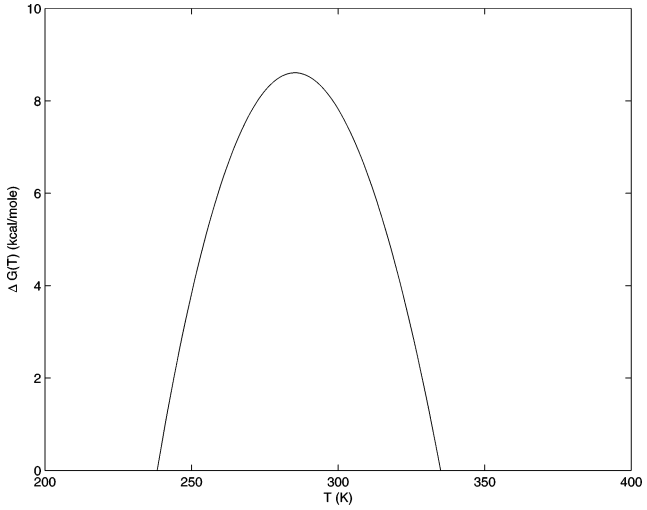

FIGURE 1: Hypothetical protein stability curve constructed by using the average values of the thermodynamic parameters (ΔH_G, ΔC_p, and T_G) for 12 single domains. The *X*-axis indicates the temperature, and the *Y*-axis indicates the $\Delta G(T)$ (the free energy difference between the denatured and native states).

thermodynamic data on proteins that show reversible two state folding ⇌ unfolding transitions around neutral pH, we have found that the majority of the proteins are maximally stable around room temperature (*9*), consistent with the hydrophobic effect being the major contributor to their stability (*5*, *10*).

Here we focus on 12 single domains which show reversible two-state folding ⇌ unfolding transitions at or near neutral pH and are maximally stable around room temperature. We have supplemented our results using protein thermodynamics simulations. We find that the temperature range (T_{Range}) over which a protein or a domain native state is thermodynamically stable is more sensitive to the variations in curvature ($-\Delta C_p/T_G$ at T_G) than to variations in the slope ($-\Delta S_G$ at T_G) of its stability curve. Since the slope and curvature of a protein stability curve depend on the protein size, T_{Range} shows a negative correlation with protein size. Protein stability curves are broader for small proteins and narrower for larger ones. The T_{Range} also shows a positive correlation with maximal protein stability [$\Delta G(T_S)$]. This correlation becomes stronger upon normalization of maximal protein stability with protein size. The melting temperature (T_G) is negatively correlated with protein size, indicating that smaller proteins have higher melting temperatures. The hydrophobic effect in single domains helps rationalize these observations. Our study has some interesting implications for protein thermostability. The genomic scale observation that thermophilic proteins are shorter, largely via deletion and/or shortening of loops (*11*), may be related to this effect. Our study suggests why it is advantageous for the protein structure to differentiate into multiple hydrophobic folding units and domains rather than exist as a large hydrophobic globule in water.

MATERIALS AND METHODS

Proteins in Our Study and Calculation of Derived Parameters. We have already described the procedure for data collection and quality control (*9*). Briefly, our data set consists of 31 proteins with two-state folding ⇌ unfolding transitions in the pH range of 6−8, with ≥90% reversibility. The cooperativity ratio, *R*, for these proteins lies in the range of 0.9−1.10. Of the 31 proteins, 26 do not contain a homologue in our data set. Of the 26 unique proteins, 20 are maximally stable around room temperature with the temperature of maximal stability (T_S) being 20 ± 8 °C. Thirteen of these 20 are single domains. Table 1 summarizes the experimental data collected from the literature on these 13 proteins. The proteins are 7 kDa DNA binding protein from *Sulfolobus solfataricus* [Sso7d (*12*)], *Thermotoga maritima* cold shock protein [TmCsp (*13*)], *Bacillus subtilis* histidine phosphocarrier protein [BsHPr (*14*, *15*)], λ repressor$_{6-85}$ (*16*), Barstar (*17*), staphylococcal nuclease [Snase (*18*)], *Escherichia coli* ferricytochrome b_{562} [FeCyt b_{562} (*19*)], *T. maritima* glutamate dehydrogenase monomer domain II [GDH domain II (*20*)], residues 109−212 in the C_L fragment of immunoglobulin Igλ [C_L of Igλ (*21*)], Kunitz type soybean trypsin inhibitor [K-STI (*22*)], *Anabaena* apoflavodoxin (*23*), *E. coli* thioredoxin (*24*), and the activation domain of human procarboxypeptidase A2 [ADA2h (*25*)]. In addition to these single-domain proteins, we have also utilized the thermodynamic data for homologues of Sso7d and BsHPr to validate our results. The homologues of Sso7d are Sac7d (*26*) and Bruton's tyrosine kinase [Btk (*27*)]. For BsHPr, we have used its homologue, *E. coli* histidine phosphocarrier protein [EcHPr (*28*)].

The three-dimensional (3D) structures were taken from the Protein Data Bank (*29*), if available. The PDBsum[2] database was used to obtain the CATH domain assignments. We have computed several structural parameters for the proteins. The Access_surf program in the Protstat group of the Homology module in InsightII was used to compute the total (ASA_{tot}), polar (ASA_{pol}), and nonpolar (ASA_{nonpol}) accessible surface areas. The hydrophobicity of a protein is computed as the ratio of the buried nonpolar surface area to the total nonpolar surface area (*30*, *31*). Compactness is defined as the ratio of the volume of the protein to the volume of the sphere that has the same surface area as the protein (*30*, *31*). The average occluded surface parameter (OSP) values were calculated using the OS71 program available at http://www.csb.yale.edu. This parameter is a measure of the average atomic packing in the proteins (*32*, *33*).

Along with the experimental thermodynamic parameters, the melting temperature (T_G), the enthalpy change at T_G (ΔH_G), and the heat capacity change (ΔC_p) taken from the literature (Table 1), we have computed additional thermodynamic parameters using the Gibbs−Helmholtz equation (*1*, *2*).

The slope of the protein stability curve at T_G is given by

$$-\Delta S_G = -\Delta H_G/T_G \qquad (2)$$

The curvature of the protein stability curve at T_G is $-\Delta C_p/T_G$.

The temperature at which the protein is maximally stable (T_S) is given by

$$T_S = T_G \exp[-\Delta H_G/(T_G \Delta C_p)] \qquad (3)$$

The maximal protein stability [$\Delta G(T_S)$], the maximum height

[2] At http://www.biochem.ucl.ac.uk/bsm/pdbsum/index.html.

Table 1: Data for 13 Unique Single Domains in Our Study[a]

protein	N_{res}	PDB entry	method	resolution (Å)	T_G (°C)	ΔH_G (kcal/mol)	ΔC_p (kcal mol^{-1} K^{-1})
Sso7d	62	1SSO	NMR		97.8	63.4	0.62 ± 0.04
TmCsp	66				82	62.6	1.1 ± 0.1
BsHpr	87	2HID	NMR		73.4 ± 0.2	58.1 ± 1.7	1.17 ± 0.05
λ repressor$_{6-85}$	80	1LMB	X-ray	1.7	57.2 ± 0.1	68.0 ± 1.0	1.44 ± 0.03
Barstar	90	1A19	X-ray	2.76	72.7 ± 0.3	72.4 ± 3.6	1.27 ± 0.24
Snase	149	1EY0	X-ray	1.6	52.8	96 ± 2	2.2
FeCyt b_{562}	106	1QPU	NMR		67.2 ± 0.5	94 ± 5	2.4 ± 0.4
GDH domain II	150	1B26	X-ray	3.0	69.6	70.2 ± 4.0	1.4 ± 0.3
C_L of Igλ	104	1A8J	X-ray	2.7	61 ± 0.5	66.5 ± 0.9	1.8
K-STI	181	1AVU	X-ray	2.3	59.0	102.5 ± 1.4	2.6 ± 0.1
apoflavodoxin	168	1FTG	X-ray	2.0	57.3 ± 0.1	63.1 ± 0.7	1.34 ± 0.02
thioredoxin	108	2TRX	X-ray	1.68	87.0	106.9 ± 1.1	1.66 ± 0.05
ADA2h	80	1AYE	X-ray	1.8	77.0	47.6	0.86 ± 0.33

[a] All proteins show highly cooperative and reversible two-state folding ⇌ unfolding transitions at or near neutral pH and are maximally stable around room temperature. The full names of these proteins are given in Materials and Methods. Data were taken from ref *9* and from published literature.

of the protein stability curve, is given by

$$\Delta G(T_S) = \Delta H_G - (T_G - T_S)\Delta C_p \quad (4)$$

The temperature range (T_{Range}) over which the native state of the protein is thermodynamically stable is given by

$$T_{Range} = T_G - T_G' \quad (5)$$

Additionally, the cold denaturation temperature (T_G') and the protein stability at room temperature [ΔG(RT)] were extrapolated from the protein stability curves. Cold denaturation has not been experimentally demonstrated for any protein in our data set (see refs *12*−*28* for the experimental details for the proteins in this study). In our study, $\Delta G(T_S)$ and ΔG(RT) have similar values since the proteins are maximally stable around room temperature.

Determination of Linear Correlation Coefficients among Various Parameter Pairs. Each pair of structural and thermodynamic parameters (x, y) for proteins in our study was fitted with a least-squares line. The linear correlation coefficient for the parameter pair is calculated by (*34*)

$$r = [n\sum xy - (\sum x)(\sum y)]/\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]} \quad (6)$$

where n is the number of proteins in our study.

Would the Correlations Observed in Our Data Sets Also Hold for Proteins in General? The observed correlations among the parameter pairs in a sample of proteins may (or may not) be observed for proteins in general or in other protein samples. The sampling theory was used to determine whether the correlations observed in our study are also relevant to proteins in general and to other protein samples. A null hypothesis was formulated as follows. The protein population correlation coefficient (ρ) for a given parameter pair is zero (H_o, $\rho = 0$), while the linear correlation coefficient for the same parameter pair is $r(\neq 0)$ in our study. The null hypothesis was tested by computing the t value (*34*):

$$t = r\sqrt{n-2}/\sqrt{1-r^2} \quad (7)$$

We reject the null hypothesis at a 95% ($p < 5 \times 10^{-2}$) or 99.5% ($p < 5 \times 10^{-3}$) level of confidence if the computed t value is greater than $t_{0.95}$ or $t_{0.995}$, respectively, for n proteins (*34*). Rejection of the null hypothesis for a parameter pair indicates that the two parameters are likely to be correlated with each other in protein populations and other samples from these populations are likely to show these correlations. Table 2 provides the threshold t values and linear correlation coefficients r for various numbers of proteins in our study for which linear regressions were performed.

In addition to the above statistical tests, the linear correlation coefficients were recomputed upon substitution of the proteins in our data set with their homologues, where such data are available. The substitutions were made both

Table 2: Thresholds for r and t Values at Various Levels of Confidence[a]

	95% level of confidence		99.5% level of confidence	
no. of proteins	$t_{0.95}$	$\lvert r_{0.95}\rvert$	$t_{0.995}$	$\lvert r_{0.995}\rvert$
13	1.77	0.47	3.01	0.67
12	1.78	0.49	3.06	0.70
11	1.80	0.52	3.11	0.72
8	1.86	0.60	3.36	0.81

[a] Given are the thresholds for t values and linear correlation coefficients (r) for rejecting the null hypothesis at 95% and 99.5% levels of confidence for different numbers of proteins.

Table 3: Average Values and Ranges for the Thermodynamic Parameters for Proteins in Our Study[a]

T_G (K)	343 ± 14	325−371
ΔH_G (kcal/mol)	77 ± 18	58−107
ΔC_p (kcal mol^{-1} K^{-1})	1.6 ± 0.6	0.6−2.6
$-\Delta S_G$ (cal mol^{-1} K^{-1})	−225 ± 53	−167 to −309
$-\Delta C_p/T_G$ (cal mol^{-1} K^{-1})	−4.7 ± 1.8	−7.8 to −1.7
T_G' (K)	248 ± 17	200−268
T_S (K)	294 ± 8	281−304
$\Delta G(T_S)$ (kcal/mol)	5.5 ± 1.6	3.5−9.0
ΔG(RT) (kcal/mol)	5.3 ± 1.6	3.5−9.0
T_{Range} (K)	95 ± 27	68−171
N_{res}	113 ± 40	62−181

[a] Given are the average values and ranges of the thermodynamic parameters for single domains in our data set. One protein, ADA2h, was not included because of large uncertainty in its ΔC_p value. The values of the thermodynamic parameters, T_G, ΔH_G, and ΔC_p, for individual proteins were taken from published experimental data. The values for the other parameters were derived as described in Materials and Methods. The average number of residues (protein size) is also presented.

singly (replacement of one protein with its homologue) and in combinations. We also repeated the calculations by arbitrarily adding and removing a protein from the data set.

Protein Thermodynamics Simulations. We have used experimental thermodynamic data for 12 single-domain two-state proteins to guide simulations of protein thermodynamics. The simulations use the Gibbs−Helmholtz equation (eq 1) to explore the consequences of systematic variations in the thermodynamic parameters, ΔH_G, ΔC_p, and T_G. The variations were performed within the observed ranges of these parameters for the 12 single-domain two-state proteins (Table 3). The procedure followed here is similar to the grid search/conformational sampling method used in molecular simulations. Two different simulation experiments were performed. In both, ΔH_G and ΔC_p were varied over the ranges of 50−150 kcal/mol and 0.5−3.0 kcal mol^{-1} K^{-1}, respectively. ΔH_G was varied in steps of 0.1 kcal/mol and ΔC_p in steps of 0.01 kcal mol^{-1} K^{-1}. We have selected those sets of thermodynamic parameters (ΔH_G, ΔC_p, and T_G) which yield a temperature of maximal stability of 298 K (eq 3).

In the first simulation experiment, we selected the thermodynamic parameter sets (ΔH_G, ΔC_p, and T_G) which yield a constant maximal stability [$\Delta G(T_S) = 5.5$ kcal/mol; stability curves plotted in Figure 2a], a constant slope at T_G ($-\Delta S_G = -225$ cal mol^{-1} K^{-1}; stability curves plotted in Figure 2b), and a constant curvature at T_G ($-\Delta C_p/T_G = -4.7$ cal mol^{-1} K^{-1}; stability curves plotted in Figure 2c). In these calculations, T_G was allowed to vary in the range of 300−500 K. The selected thermodynamic parameter sets were used to plot hypothetical protein stability curves using the Gibbs−Helmholtz equation. The purpose of this experiment was to probe the sensitivity of T_{Range} to variations in the slope and curvature of the protein stability curves.

In the second experiment, T_G was varied in the range of 300−400 K in 5 K steps along with the variations in ΔH_G and ΔC_p stated above. The averages of the selected thermodynamic parameters were computed and used to plot the protein stability curves (shown in Figure 4). This experiment addresses the question of how proteins achieve higher melting temperatures.

RESULTS

Our data set consists of 13 unique single domains that are maximally stable around room temperature. Each domain contains a single hydrophobic core, and the Gibbs−Helmholtz equation applies best to this class. In the calculations presented here, we have not included the single-domain ADA2h (*25*). ADA2h has a rather large uncertainty in its ΔC_p value (Table 1) that affects its protein stability curve and the accuracy of the derived thermodynamic parameters (*9*). Table 3 presents averages and ranges of variation seen in the thermodynamic parameters for the remaining 12 single domains in the data set. The 12 single domains have relatively small average values for the enthalpy change at the melting temperature ($\Delta H_G = 77 \pm 18$ kcal/mol) and the heat capacity change ($\Delta C_p = 1.6 \pm 0.6$ kcal mol^{-1} K^{-1}). The maximal stabilities of these proteins are less than 10 kcal/mol [average $\Delta G(T_S) = 5.5 \pm 1.6$ kcal/mol], and their melting temperatures span the range typically seen for proteins from the mesophilic and thermophilic organisms (*35*). The native states of these proteins are stable over a temperature range of ~100 K (average $T_{Range} = 95 \pm 27$ K). Figure 1 shows a hypothetical protein stability curve plotted using average thermodynamic parameters of the 12 single domains.

Correlations among Various Parameters for 12 Single Domains

Table 4 presents the linear correlation coefficients (r) among the various thermodynamic parameters for the 12 single domains. The protein stability curve (Figure 1) indicates that the temperature range over which the native state of a single-domain protein is thermodynamically favored ($T_{Range} = T_G - T_G'$) should depend on the slope ($-\Delta S_G$ at T_G) and curvature ($-\Delta C_p/T_G$ at T_G) of the protein stability curve. The temperature of maximal stability (T_S) and maximal protein stability [$\Delta G(T_S)$] also affect the T_{Range}. In our data set, the T_{Range} shows significant correlations with the curvature and maximal protein stability for 12 single domains. The correlation between T_{Range} and the curvature ($r = 0.74$) is significant at the 99.5% level of confidence. Furthermore, T_G and T_G' (estimated cold denaturation temperature) are oppositely correlated with $-\Delta C_p/T_G$ (Table 4) and with ΔC_p. A similar behavior of T_G, T_G', and T_{Range} is also seen with respect to maximal protein stability. The correlations among these parameters become highly significant (99.5% level of confidence) when $\Delta G(T_S)$ is normalized with respect to the protein size [$\Delta G(T_S) = \Delta G(T_S)/N_{res}$]. These observations indicate that these parameters may be correlated for reversible two-state single-domain proteins in general.

The T_{Range} does not show significant correlations with either the slope or T_S. In our data set, T_S is restricted to a small range near room temperature and both T_G and T_G' show positive correlations with T_S. Because the slope of a protein stability curve at T_G is also related to the enthalpy change at T_G ($-\Delta S_G = -\Delta H_G/T_G$), we find that T_{Range} does not show a significant correlation with ΔH_G either. However, the correlation between T_{Range} and Δh_G ($=\Delta H_G/N_{res}$) is significant at the 95% level of confidence (Table 4). We note that maximal protein stability also depends on both slope and curvature of the stability curve (eq 4). Furthermore, the correlation between the slope and curvature is highly significant (99.5% level of confidence) for 12 single domains ($r = 0.83$). These observations indicate that there may still be a significant contribution from the slope of the protein stability curve to its T_{Range}. By using the least-squares method, we have derived a parametric equation that describes the relationship among the slope, curvature, and T_{Range} for the 12 single-domain proteins. This equation is

$$T_{Range}\,(\mathrm{K}) = -0.45 \times \text{slope}\,(\text{cal mole}^{-1}\,\mathrm{K}^{-1}) + 22.46 \times \text{curvature}\,(\text{cal mol}^{-1}\,\mathrm{K}^{-1}) + 97.96$$

The root-mean-square error for this equation is 12 K.

Taken together, these results indicate that T_{Range} may be more sensitive to variations in the curvature rather than to the slope of the protein stability curve. To gain further insight into this issue, we have used a grid search-based simulation experiment to obtain sets of the three thermodynamic parameters (ΔH_G, ΔC_p, and T_G) in such a way that they yield

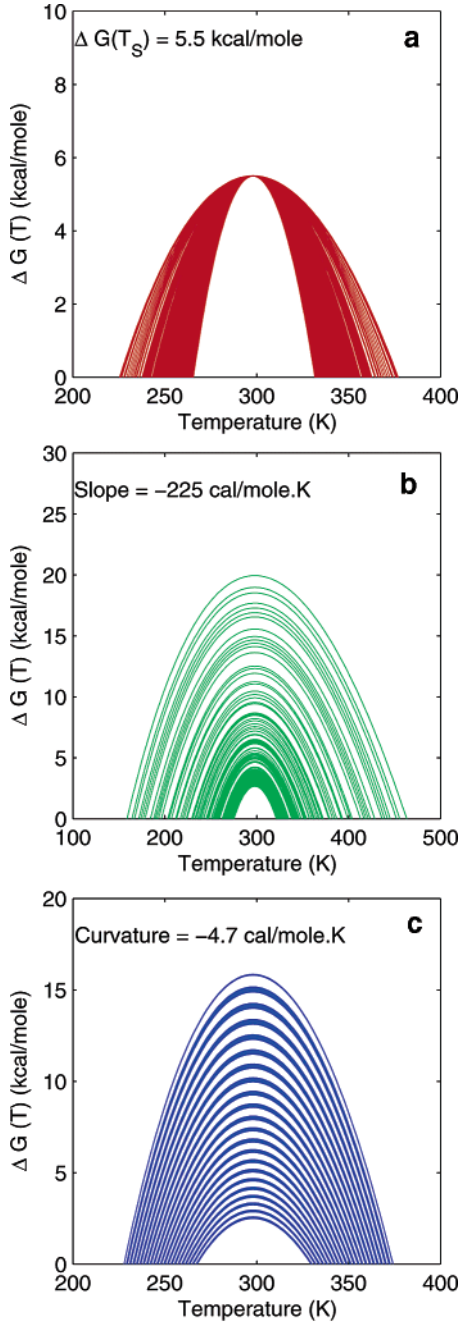

FIGURE 2: Results from simulations of the protein thermodynamic data. The values of ΔH_G, ΔC_p, and T_G were varied systematically in small steps over a range inspired by the experimentally determined values of these parameters for 12 single domains. In each panel, the protein stability curves for selected combinations of these parameters are plotted. (a) Protein stability curves for 211 sets of the thermodynamic parameters which yield a constant maximal protein stability [$\Delta G(T_S) = 5.5$ kcal/mol]. (b) Protein stability curves for 96 sets of thermodynamic parameters with a constant slope ($-\Delta S_G = -225$ cal mol^{-1} K^{-1}). (c) Protein stability curves for 312 sets of the thermodynamic parameters with a constant curvature ($-\Delta C_p/T_G = -4.7$ cal mol^{-1} K^{-1}). The constant values of maximal stability, slope, and curvature are the average values of these parameters for 12 single domains (Table 3), and all the curves show maximal stability at 298 K. This figure indicates that the temperature range (T_{Range}) of thermodynamic stability is more sensitive to variations in curvature than to those in slope.

the temperature of maximal stability at room temperature (298 K). The ranges for the thermodynamic parameters ΔH_G and ΔC_p were inspired by the variations in the values of these parameters shown by 12 single domains (Table 3). The hypothetical protein stability curves at constant values of maximal stability, curvature, and slope were plotted using these selected sets (Figure 2). A total of 211 parameter sets yield a maximal stability [$\Delta G(T_S)$] of 5.5 kcal/mol. Ninety-six sets have a slope at T_G ($-\Delta S_G$) of -225 cal mol^{-1} K^{-1}, and 312 sets have a curvature at T_G ($-\Delta C_p/T_G$) of -4.7 cal mol^{-1} K^{-1}. The selected constant values of maximal stability, slope, and curvature at T_G are the average values of these parameters for the 12 single domains. This experiment shows that different sets of slope and curvature values can yield the same maximal protein stability (Figure 2a). The stability curves exhibit a wider range when the slope is kept constant and the curvature is varied (Figure 2b) rather than *vice versa* (Figure 2c). These again indicate the greater sensitivity of T_{Range} to variations in curvature. An analogous experiment that allowed the temperature of maximal stability (T_S) to be varied in the range of 273−310 K yielded similar results (data not shown).

Among the 12 single-domain proteins, both T_G and T_{Range} are negatively correlated with protein size. The slope and curvature also show significant correlations with N_{res}. These correlations are significant at the 95% level of confidence. Taken together, we obtain a consistent picture with respect to the single-domain proteins. The stability curves of larger proteins have greater slopes and curvatures. This leads to sharper and narrower curves, resulting in smaller temperature ranges over which their native states are thermodynamically stable. We chose a pictorial example to illustrate this point. Figure 3 compares the sizes and stability curves of a small (Sso7d, 62 residues, $T_{Range} = 170$ K), a midsized (Barstar, 90 residues, $T_{Range} = 102$ K), and a large (K-STI, 181 residues, $T_{Range} = 73$ K) single-domain protein.

Additionally, ΔC_p, ΔS_G, and ΔH_G also show strong correlations among themselves and with $\Delta G(T_S)$. We have normalized these parameters and ΔG(RT) by the number of residues (Table 4). For the 12 single domains, the normalization yields negative correlations of these with N_{res}. There are strong correlations among T_G and $\Delta g(T_S)$, Δh_G and $\Delta g(T_S)$, and Δs_G and $\Delta g(T_S)$, being significant at the 99.5% level of confidence (*9*). Thus, the average enthalpic, entropic, and free energy contributions decrease with the increase in protein size, consistent with an upper limit for maximal protein stability. We also notice a strong negative correlation between T_G' and $\Delta g(T_S)$ (Table 4). Combined with T_G versus $\Delta g(T_S)$, this results in a strong positive correlation between T_{Range} and $\Delta g(T_S)$. Table 5 gives the least-squares regression line equations for the five important parameter pairs, N_{res} and T_G, N_{res} and T_{Range}, ΔS_G and ΔC_p, ΔC_p and T_{Range}, and $\Delta g(T_S)$ and T_{Range}, for the 12 single domains.

Of 12 single domains in our data set, the thermodynamic parameters for the homologues of two proteins, namely, Sso7d and BsHPr (Table 1), are also available (see Table 1 of ref *9*). We have substituted the homologous proteins and recalculated the linear correlation coefficients. The following protein exchanges were made: Sso7d → Sac7d, Sso7d → Btk, BsHPr → EcHPr, Sso7d → Sac7d and BsHPr → EcHPr, and Sso7d → Btk and BsHPr → EcHPr (see Table 6 for details). The results have not changed appreciably. Arbitrary removal and addition of a protein to our data set also do not affect the results (Table 7). These tests give us confidence that despite the limitation of our small data set, it is very

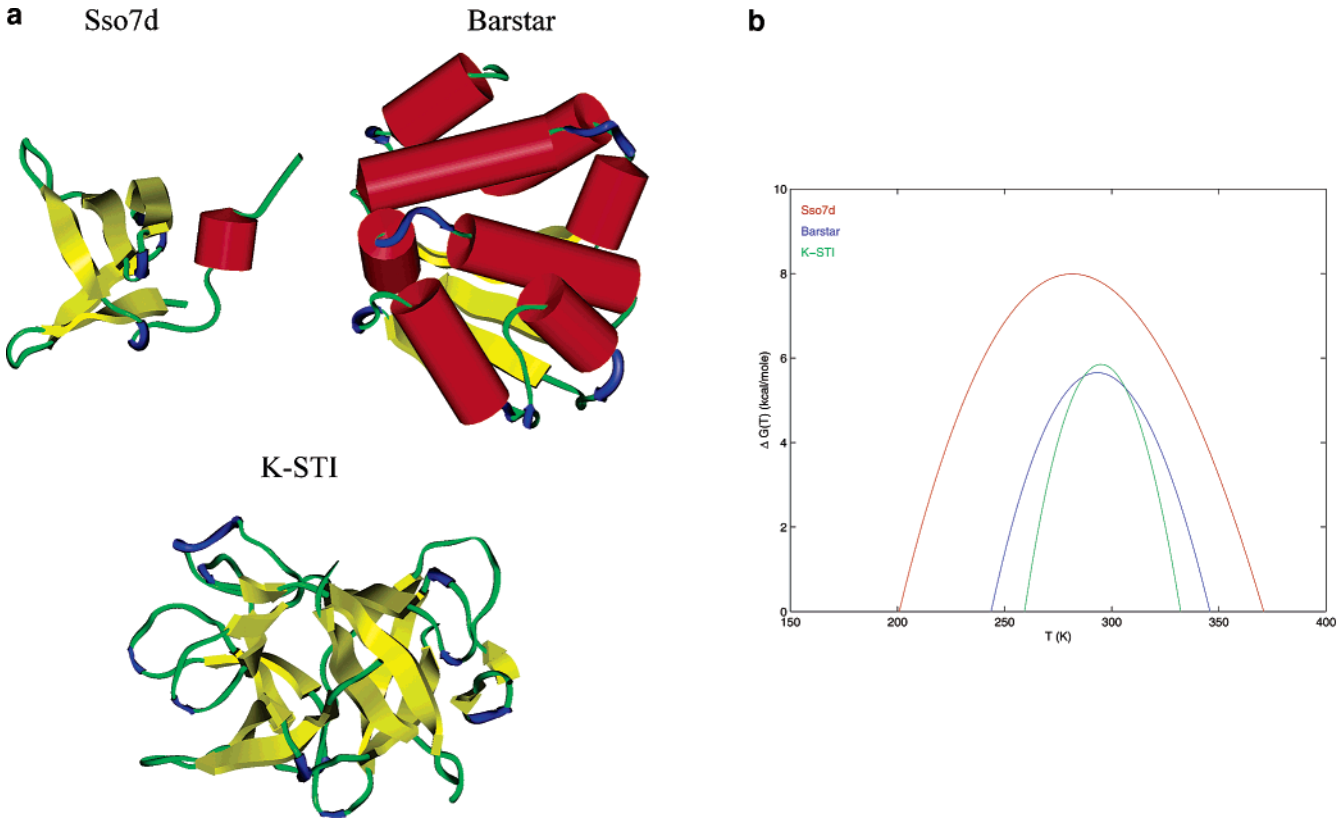

FIGURE 3: Example of the relationship between the protein size and temperature range of its thermodynamic stability. We illustrate our point using three single-domain proteins: a small 62-residue hyperthermophilic protein, Sso7d (PDB entry 1SSO), a midsized 90-residue Barstar (PDB entry 1A19), and a large 181-residue K-STI (PDB entry 1AVU). Sso7d is the smallest single-domain protein in our data set, and K-STI is the largest. The size of Barstar lies closer to the average protein size in 12 single domains. (a) 3D structures of the three proteins plotted using the SecondaryRender program of the Molecule module in InsightII version 2000. (b) The plot compares the stability curves for the three proteins shown in panel a. Sso7d (in red) has the largest and K-STI (green) the smallest temperature range (T_{Range}) of thermodynamic stability in the plot. That of Barstar (blue) lies between these two.

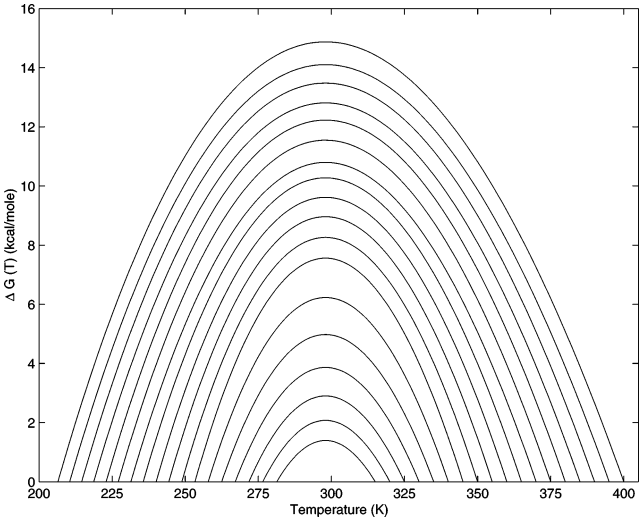

FIGURE 4: Protein stability curves plotted every 5 K in the melting temperature range of 300−400 K. The curves were plotted using the averages of the data collected from simulations of the experimentally determined ΔH_G, ΔC_p, and T_G for 12 single domains. A total of 8686 sets of thermodynamic parameters were used to compute the averages to plot the protein stability curves. This figure indicates that proteins need to upshift and broaden their stability curves to achieve higher melting temperatures.

likely that the correlations observed here would also apply to the reversible two-state single domains in general.

To investigate whether any protein structural property plays a role in determining the thermodynamic parameters, we have performed correlation analysis between the structural (Table 8a) and thermodynamic parameters for eight single-domain monomers for which crystallographic data are available. We find that parameters related to atomic packing (compactness and average occluded surface area) are not correlated with the thermodynamic parameters. Hydrophobicity, the proportion of nonpolar buried surface area, is also uncorrelated. T_G and $\Delta G(T_S)$ do not show any significant correlation with parameters based on surface areas or on the proportions of charged, aromatic, or nonpolar residues. At the same time, the ASA magnitudes are correlated with parameters which determine the slope and curvature of the protein stability curves (Table 9). The nonpolar accessible surface area is positively correlated with protein size and ΔC_p. Similar observations have been made by Makhatadze and Privalov (*36*) in a different data set of single-domain proteins. The correlation between the nonpolar surface area and ΔC_p has been described (*10*, *37*, *38*).

Implications of the Observed Correlations

The observed correlations in the 12 single domains have interesting implications for protein structure and thermostability. Below, we discuss some of these in some detail.

Temperature Range and the Size of Protein Domains. Most proteins consist of structural building blocks, hydrophobic folding units, domains, and subunits. Our results imply that there may be an upper limit to the size of a single cooperative hydrophobic globule. As the size grows, the temperature

Table 4: Correlations among Various Thermodynamic Parameters in 12 Single Domains[a]

				Un-Normalized Parameters					
parameter	T_G	ΔH_G	ΔC_p	$-\Delta S_G$	T_G'	T_S	$\Delta G(T_S)$	T_{Range}	$-\Delta C_p/T_G$
N_{res}	−0.63		0.62	0.54				−0.53	−0.64
T_G			−0.63		0.58	(**0.74**)	0.60	**0.86**	0.67
ΔH_G			**0.78**	**0.99**			0.56 (**0.81**)		**−0.75**
ΔC_p				**0.85**	−0.66			**−0.73**	**−1.00**
$-\Delta S_G$							(**−0.73**)		**0.83**
T_G'						**0.72**	−0.58	**−0.91**	−0.65
T_S									
$\Delta G(T_S)$								0.66	
T_{Range}									**0.74**

				Normalized Parameters					
parameter pair	T_G	Δh_G	ΔC_p	$-\Delta S_G$	T_G'	T_S	$\Delta g(T_S)$	T_{Range}	$-\Delta C_p/T_G$
N_{res}	−0.63	**−0.80**		**0.79**			**−0.71**	−0.53	
T_G		0.69		−0.59			**0.86**	**0.86**	
Δh_G			(**0.74**)	**−0.99**			**0.84** (**0.92**)	0.58	
ΔC_p				−0.57 (**−0.79**)	0.54	0.51			**−0.99**
$-\Delta S_G$							**−0.78** (**−0.88**)		0.49
T_G'						**0.72**	**−0.79**	**−0.91**	−0.58
T_S									
$\Delta g(T_S)$								**0.92**	
T_{Range}									0.50

[a] Linear correlation coefficients (r) among the thermodynamic parameters in 12 single-domain proteins in our study. The correlations which are significant at the ≥95% level of confidence are shown (bold if significant at the 99.5% level). The cutoff values for r were taken from Table 2. A hyperthermophilic single-domain protein, Sso7d, appears to be an outlier in plots of a few parameter pairs. When the calculations were performed without Sso7d, the results were qualitatively identical. The r values in parentheses are for the remaining 11 single-domain proteins, shown only when the correlation improves.

Table 5: Regression Line Equations for Parameter Pairs in 12 Single Domains[a]

parameter pair	regression line equation	r	root-mean-square error
N_{res}, T_G	T_G (K) = $-0.21N_{res} + 367.09$	−0.63	10.18 K
N_{res}, T_{Range}	T_{Range} (K) = $-0.36N_{res} + 135.21$	−0.53	22.30 K
ΔC_p, T_{Range}	T_{Range} (K) = $-34.34\Delta C_p$ (kcal mol^{-1} K^{-1}) + 149.0	−0.73	18.06 K
ΔS_G, ΔC_p	ΔC_p (cal mol^{-1} K^{-1}) = $9.3\Delta S_G$ (cal mol^{-1} K^{-1}) − 508.6	0.85	288.1 cal mol^{-1} K^{-1}
$\Delta g(T_S)$, T_{Range}	T_{Range} (K) = $0.87\Delta g(T_S)$ (kcal/mol) + 46.33	0.92	10.15 K

[a] Regression line equations ($y = mx + c$) for the plots of a few of the parameter pairs in our study. r denotes the linear correlation coefficient. The lines were fitted using the least-squares method. The root-mean-square error for each equation is the root-mean-square deviation between the actual data and those predicted from the least-squares line. Units for each parameter in the equation are given in parentheses.

Table 6: Effect of Substitution of Homologous Proteins in the Data Set on Observed Correlations[a]

parameter pair	12 single domains	Sso7d → Sac7d	Sso7d → Btk	BsHPr → EcHPr	Sso7d → Btk BsHPr → EcHPr	Sso7d → Sac7d BsHPr → EcHPr
T_{Range}, $-\Delta C_p/T_G$	**0.74**	**0.75**	**0.75**	**0.77**	**0.77**	**0.78**
T_{Range}, $-\Delta S_G$						
T_{Range}, $\Delta g(T_S)$	**0.92**	**0.81**	**0.75**	**0.92**	**0.74**	**0.81**
T_{Range}, N_{res}	−0.53	−0.51	−0.50	−0.53	−0.50	−0.52
T_G, N_{res}	−0.63	−0.62	−0.59	−0.58	−0.53	−0.56
$-\Delta C_p/T_G$, N_{res}	−0.64	−0.62	−0.63	−0.62	−0.61	−0.60
-Δs_G, N_{res}	**0.79**	**0.76**	0.68	**0.85**	**0.74**	**0.82**

[a] Thermodynamic parameters are available for the homologues of two proteins in our data set. These proteins are Sso7d [homologues, Sac7d (*26*) and Btk (*27*)] and BsHPr [homologue, EcHPr (*28*)]. The thermodynamic parameters for the homologues were taken from Table 1 of ref *9*. The thermodynamic parameters were substituted both singly (e.g., Sso7d → Sac7d) and in combinations (e.g., Sso7d → Sac7d, BsHPr → EcHPr). The linear correlation coefficients were recomputed for all parameter pairs shown in Table 4 upon each substitution. Here, we show only those parameter pairs which bear on important results of this study. $-\Delta s_G$ and $\Delta g(T_S)$ were obtained by normalizing $-\Delta S_G$ and $\Delta G(T_S)$, respectively, with the protein size (N_{res}). As in Table 4, all the linear correlation coefficients shown here are significant at the ≥95% level (bold if significant at the 99.5% level) of confidence. Qualitatively, our results remain the same upon homologue substitutions.

range over which it is thermodynamically stable appears to shrink and its melting temperature appears to drop. Extrapolations from the regression line equations in Table 5 indicate that a single hydrophobic globule of approximately 320 residues may melt around room temperature and be stable over a temperature range of only ∼20 K. This suggests a thermodynamic advantage for proteins that differentiate into multiple units (domains). Consistently, the frequency of the occurrence of single-domain proteins decreases with the increase in the protein size (*39*). A recent database analysis of protein structural domains in the Protein Data Bank by the Barton group (*40*) indicates that the majority of the protein domains contain 50−150 residues and most of these are smaller than 300 residues (Figures 2 and 4a in ref *40*). Furthermore, the majority of protein chains with >200 residues contain multiple domains with the number of

Table 7: Effect of Arbitrary Addition and Removal of a Protein in the Data Set on Observed Correlations[a]

parameter pair	11 two-state single domains[b]	12 two-state single domains	13 two-state single domains[c]
T_{Range}, $-\Delta C_p/T_G$	Y	Y	Y
T_{Range}, $-\Delta S_G$	N	N	N
T_{Range}, $\Delta g(T_S)$	Y	Y	Y
T_{Range}, N_{res}	N	Y	Y
T_G, N_{res}	Y	Y	Y
$-\Delta C_p/T_G$, N_{res}	Y	Y	Y
$-\Delta s_G$, N_{res}	Y	Y	Y

[a] The parameter pairs that show statistically significant correlations in different data sets in our study are shown. Only those parameter pairs which are important for this study are shown. $-\Delta s_G$ and $\Delta g(T_S)$ were obtained by normalizing $-\Delta S_G$ and $\Delta G(T_S)$, respectively, with the protein size (N_{res}). The symbol Y indicates that the parameter pairs show a correlation significant at the ≥95% level of confidence, indicating that this parameter pair is likely to be correlated in two-state single-domain proteins in general. The symbol N indicates that the correlation shown by the parameter pair is not significant at the 95% level of confidence. [b] The data set of 11 two-state single domains was obtained by removing the hyperthermophilic single-domain protein Sso7d (*12*) from the data set of 12 two-state single domains, the focus of this study. [c] The data set of 13 two-state single domains was obtained by adding the activation domain of human procarboxypeptidase A2 [ADA2h (*25*)] to the 12 two-state single domains. Note that ADA2h was not included in the original calculations because of the rather large uncertainty in its ΔC_p value.

domains increasing with the size of the protein chains (*40*). Previously, Xu and Nussinov (*41*) have used an empirical function that describes the energetic balance between the gain in enthalpy due to favorable internal interactions and the loss of conformational entropy in the folded state to argue that a protein chain should be neither too short nor too long to form a stable single domain. Their study indicates that the optimal size for a single domain is 100 residues. Consistently, the average size of the 12 single domains in our database is 113 ± 40 residues with the range being 62−181 residues.

Are Small Proteins More Thermostable? T_G, T_{Range}, and curvature are all negatively correlated with protein size (Table 4). The negative correlation between curvature and N_{res} is significant even at the 99.5% level of confidence ($r = -0.89$). These indicate that smaller proteins may melt at higher temperatures and may be stable over wider temperature ranges. In few specific instances, it was noted that small single-domain proteins have high melting temperatures (*12*, *13*, *26*, *42*). An extreme example is that of *Pyrococcus furiosus* rubredoxin ($N_{res} = 53$) whose T_G is reported to be ∼200 °C (*43*). Rubredoxin does not show a simple reversible two-state folding ⇌ unfolding transition, and its extraordinarily high melting temperature (T_G) is not predicted by our regression equation relating protein size and T_G (Table 5). Notwithstanding this minor discrepancy, our analysis suggests that, in general, small proteins may indeed be more thermostable. This is mainly due to smaller heat capacity (ΔC_p) values for such proteins. Consistently, genomic scale comparisons among the thermophilic and mesophilic proteins have shown that thermophilic proteins contain fewer residues than mesophilic proteins, largely the outcome of loop shortening and/or deletion (*11*).

Upshift and Broadening of Protein Stability Curves Leads to Greater Thermostability. A higher melting (transition) temperature for a thermophilic protein can be attained in one of three ways (*26*, *42*, *44*−*46*): The stability curve of the thermophilic protein can be upshifted as compared to those of its mesophilic homologue(s), the curvatures may be different, or there may be a left or right shift of the curve. By comparing the stability curves of 19 homologous thermophilic and mesophilic proteins in five different families, we have observed that thermophilic proteins have broader and upshifted stability curves (*47*). As a result, the temperature ranges of thermodynamic stability for the thermophilic proteins are larger than those for the mesophilic homologues (Figure 1 in ref *47*). To further explore this issue, here we again use the grid search methodology to select sets of thermodynamic parameters that show maximal stability at room temperature. In the T_G range of 300−400 K, the average values of ΔH_G and ΔC_p for selected sets were computed at 5 K T_G intervals. A total of 8686 sets of thermodynamic parameters were used for the calculation of these averages. The average thermodynamic parameters were used to plot the hypothetical protein stability curves shown in Figure 4. Again, this experiment shows that higher melting temperatures are achieved via upshift and broadening of the curves. This suggests that proteins with higher T_G values will also have lower T_G' values, indicating a greater temperature range of thermodynamic stability. Varying the T_S (temperature of maximal stability) in the range of 273−310 K yielded similar results (data not shown).

DISCUSSION AND CONCLUSIONS

Protein thermodynamics are essential for understanding structure and function. One way to study protein thermodynamics is via collection and analysis of experimental data on protein folding ⇌ unfolding transitions. However, there are limitations. Thermodynamic experiments are performed in different laboratories under different experimental conditions. The folded and unfolded states of the different proteins are also different. The unfolded state may still contain residual structure, sometimes with native-like topology (*48*). The errors in experimental estimates of thermodynamic parameters can be propagated to the derived parameters and make the correlation analyses difficult. Recently, Pace et al. (*7*) have illustrated that values of ΔC_p tend to be more unreliable than T_G and ΔH_G values. Experimental T_G is usually determined accurate within 1%, while ΔH_G is accurate within 5%. For most proteins in our study, the estimated errors in ΔC_p are within 10% and do not considerably affect the stability curves of the individual proteins (*9*). Here, we have employed a number of quality controls (see Materials and Methods and ref *9*) which give us confidence that the quality of our data is reasonably high. Furthermore, we have supplemented our results with simulations of protein thermodynamics. The simulations were performed using the values and ranges observed in the experimental data.

Here we have focused on the 12 single domains with reversible two-state folding ⇌ unfolding transitions. While thermodynamic parameters for many single domains have been experimentally measured, the number of those which follow our selection criteria [scan rate and concentration independence, constant ΔC_p, folding ⇌ unfolding transition at or near neutral pH, maximal stability around room temperature, reversibility of ≥90%, cooperativity ratio (R) of 0.9−1.1] is still quite small. As with all small data analyses, there is no guarantee that other data sets of reversible two-state single-domain proteins would also show

Table 8: Structural Parameters Derived from Crystal Structures of Eight Single-Domain Proteins[a]

protein	N_{res}	hydrophobicity (%)	compactness	AvOSP	ASA_{tot} (Å^2)	ASA_{pol} (Å^2)	ASA_{nonpol} (Å^2)
λ repressor$_{6-85}$	80	76	1.79	0.35 ± 0.13	4973	2272	2701
Barstar	90	80	1.43	0.36 ± 0.13	5103	2389	2714
Snase	149	79	1.72	0.36 ± 0.13	8224	3745	4479
GDH domain II	150	82	1.65	0.37 ± 0.14	7711	3572	4139
C_L of Igλ	104	82	2.02	0.32 ± 0.13	6350	2618	3732
K-STI	181	86	1.64	0.36 ± 0.14	9062	4290	4772
apoflavodoxin	168	86	1.47	0.38 ± 0.14	7701	4090	3611
thioredoxin	108	79	1.51	0.35 ± 0.14	5946	2618	3328

[a] Of the 12 single-domain proteins in our data set, 11 domains have 3D structures. Eight are crystal structures. These eight were used in our analysis. For GDH domain II, the atomic coordinates of the domain were extracted from PDB entry 1B26 which contains the structure for the full protein. N_{res} is the number of residues. Hydrophobicity, compactness, and AvOSP were calculated as described in Materials and Methods. AvOSP stands for average occluded surface parameter. It is the average of the occluded surface parameter for individual amino acid residues in a protein. For a residue, an OSP of 1 indicates that the residue is completely occluded (inaccessible to water). An OSP of 0 indicates that the residue is completely exposed. ASA_{tot} stands for the total accessible surface area of the protein. It consists of polar (ASA_{pol}) and nonpolar (ASA_{nonpol}) components.

Table 9: Linear Correlation Coefficients (r) among Structural and Thermodynamic Parameters for Eight Single-Domain Proteins[a]

parameter	N_{res}	Δh_G	ΔC_p	ΔS_G
ASA_{tot}	0.86	0.67	**0.88** (0.67)	0.81
ASA_{pol}	0.78			0.80
ASA_{nonpol}	0.78	0.65	**0.87** (0.78)	0.71

[a] The thermodynamic and structural parameters have been normalized by the number of residues (N_{res}). In the case of heat capacity change, the correlation coefficients given in parentheses are un-normalized values of heat capacity change and relevant structural parameter. Correlations significant at the $\geq$95% level of confidence are shown (bold if significant at the 99.5% level). The cutoff values for r were taken from Table 2.

the correlations observed in this study. However, the followings give us confidence that most of the results of this analysis will also apply to other single-domain proteins. We have used the *t* test to determine whether a parameter pair which is correlated in a data set of 12 single domains would also be correlated in reversible two-state single-domain proteins in general. In this study, we have reported only those correlations which pass this test at the $\geq$95% level of confidence. Furthermore, substitution with homologous proteins (Table 6) and arbitrary addition or removal of a protein (Table 7) in our data set do not significantly affect our results.

The 12 single-domain proteins contain single hydrophobic cores. This may be the reason behind the observed statistically significant correlations among their thermodynamic parameters despite the fact that each protein in our data set is unique. For some of the parameter pairs in Tables 4 and 9, similar correlations have been reported previously (*9*, *37*, *38*, *47*, *49*).

The temperature range over which a protein's native state is thermodynamically stable (T_{Range}) is sensitive to variations in the curvature of its stability curve. The curvature is governed by ΔC_p which mainly reflects the hydrophobic effect in the proteins (*2*, *9*, *37*, *47*). For the proteins in our data set, the hydrophobic effect is the major factor contributing to their stability (*9*). Consistently, the nonpolar surface area is correlated with protein size and ΔC_p for the eight single domains in our study for which crystal structures are available (Table 9). Exposure of the nonpolar surface to water drives protein denaturation at high and low temperatures (*50*). Hence, it is conceivable that the larger the single-domain nonpolar surface, the smaller the temperature range over which it is thermodynamically stable. This rationalizes the negative correlation seen between T_{Range} and protein size and its implications.

Protein thermodynamic simulations indicate that upshift and broadening of the protein stability curves lead to higher melting temperatures (Figure 4). Consistently, we have reported earlier that the thermophilic proteins have lower ΔC_p values and are stable over temperature ranges wider than those of their mesophilic homologues (*47*). We have interpreted this observation in terms of the formation of additional specific interactions in the thermophilic proteins such as salt bridges (*35*, *51*, *52*), aromatic clusters (*53*), and cation$-\pi$ interactions (*54*). Shortening the chain length via loop deletion and/or shortening can also reduce the ΔC_p and increase T_G (*11*). Loop deletion or shortening also reduces the conformational entropy at high temperatures.

To conclude, our results indicate that the temperature range of thermodynamic stability for a single domain depends on size. Small, two-state single-domain proteins have broader stability curves leading to T_G values higher than and T_G' values lower than those of the larger single-domain two-state proteins. The negative correlations of the temperature range with the heat capacity change and protein size point to the hydrophobic effect. It further suggests why protein structures often contain multiple hydrophobic folding units, domains, and subunits.

ACKNOWLEDGMENT

We thank Drs. Neeti Sinha, Gunasekaran Kannan, David Zanuy, and, in particular, Jacob V. Maizel for numerous helpful discussions.

REFERENCES

1. Becktel, W., and Schellman, J. A. (1987) Protein stability curves, *Biopolymers 26*, 1859−1877.
2. Privalov, P. L. (1990) Cold denaturation of proteins, *Crit. Rev. Biochem. Mol. Biol. 25*, 281−305.
3. Kauzmann, W. (1959) Some factors in the interpretation of protein denaturation, *Adv. Protein Chem. 14*, 1−63.
4. Dill, K. A. (1990) Dominant forces in protein folding, *Biochemistry 29*, 7133−7155.
5. Schellman, J. A. (1997) Temperature, stability and hydrophobic interaction, *Biophys. J. 73*, 2960−2964.
6. Pace, C. N., Shirley, B. A., McNutt, M., and Gajiwala, K. (1996) Forces contributing to the conformational stability of proteins, *FASEB J. 10*, 75−83.
7. Pace, C. N., Hebert, E. J., Shaw, K. L., Schell, D., Both, V., Krajcikova, D., Sevcik, J., Wilson, K. S., Dauter, Z., Hartley, R. W., and Grimsley, G. R. (1998) Conformational stability and

thermodynamics of folding of ribonucleases Sa, Sa2 and Sa3, *J. Mol. Biol. 279*, 271−286.
8. Pace, C. N. (2001) Polar Group Burial Contributes More to Protein Stability than Nonpolar Group Burial, *Biochemistry 40*, 310−313.
9. Kumar, S., Tsai, C. J., and Nussinov, R. (2002) Maximal stabilities of reversible two-state proteins, *Biochemistry 41*, 5359−5374.
10. Privalov, P. L., and Gill, S. J. (1988) Stability of protein structure and hydrophobic interaction, *Adv. Protein Chem. 39*, 191−234.
11. Thomson, M. J., and Eisenberg, D. (1999) Transproteomic evidence of a loop-deletion mechanism for enhancing protein thermostability, *J. Mol. Biol. 290*, 595−604.
12. Knapp, S., Karshikoff, A., Berndt, K. D., Christova, P., Atanasov, B., and Ladenstein, R. (1996) Thermal unfolding of the DNA-binding protein Sso7d from the hyperthermophile *Sulfolobus solfataricus*, *J. Mol. Biol. 264*, 1132−1144.
13. Wassenberg, D., Welker, C., and Jaenicke, R. (1999) Thermodynamics of the unfolding of the cold shock protein from *Thermotoga maritima*, *J. Mol. Biol. 289*, 187−193.
14. Pfeil, W. (1998) *Protein stability and folding. A collection of thermodynamic data*, Springer-Verlag, Heidelberg, Germany.
15. Scholtz, J. M. (1995) Conformational stability of HPr: The histidine-containing phosphocarrier protein from *Bacillus subtilis*, *Protein Sci. 4*, 35−43.
16. Huang, G. S., and Oas, T. G. (1996) Heat and cold denatured states of monomeric λ repressor are thermodynamically and chemically equivalent, *Biochemistry 35*, 6173−6180.
17. Schoppe, 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.
18. Shortle, D., Meeker, A. K., and Freire, E. (1988) Stability mutants of staphylococcal nuclease: Large compensating enthalpy−entropy changes for the reversible denaturation reaction, *Biochemistry 27*, 4761−4768.
19. Feng, Y., and Sligar, S. G. (1991) Effect of heme binding on the structure and stability of *Escherichia coli* apocytochrome b_{562}, *Biochemistry 30*, 10150−10155.
20. Consalvi, V., Chiaraluce, R., Giangiacomo, L., Sandurra, R., Christova, P., Karshikoff, A., Knapp, S., and Ladenstein, R. (2000). Thermal unfolding and conformational stability of recombinant domain II of glutamate dehydrogenase from the hyperthermophile *Thermotoga maritima*, *Protein Eng. 13*, 501−507.
21. Okajima, T., Kawata, Y., and Hamaguchi, K. (1990) Chemical modification of tryptophan residues and stability changes in proteins, *Biochemistry 29*, 9168−9175.
22. Fukada, H., Kitamura, S., and Takahashi, K. (1995) Calorimetric study of the thermal unfolding of Kunitz-type soybean trypsin inhibitor at pH 7.0, *Thermochim. Acta 266*, 365−372.
23. Genzor, C. G., Beldarrain, A., Gomez-Moreno, C., Lopez-Lacomba, J. L., Cortijo, M., and Sancho, J. (1996) Conformational stability of apoflavodoxin, *Protein Sci. 5*, 1376−1388.
24. Santoro, M. M., and Bolen, D. W. (1992) A test of the linear extrapolation of unfolding free energy changes over an extended denaturant concentration range, *Biochemistry 31*, 4901−4907.
25. Villegas, V., Azuaga, A., Catasus, L., Reverter, D., Mateo, P. L., Avilles, F. X., and Serrano, L. (1995) Evidence for a two state transition in the folding process of the activation domain of human procarboxypeptidase A2, *Biochemistry 34*, 15105−15110.
26. McCrary, B. S., Edmondson, S. P., and Shriver, J. W. (1996) Hyperthermophile protein folding thermodynamics: Differential scanning calorimetry and chemical denaturation of Sac7d, *J. Mol. Biol. 264*, 784−805.
27. Knapp, S., Mattson, P. T., Christova, P., Berndt, K. D., Karshikoff, A., Vihinen, M., Smith, C. I. E., and Ladenstein, R. (1998) Thermal unfolding of small proteins with SH3 domain folding pattern, *Proteins 31*, 309−319.
28. Nicholson, E. M., and Scholtz, J. M. (1996) Conformational stability of *Escherichia coli* HPr protein: Test of the linear extrapolation method and a thermodynamic characterization of cold denaturation, *Biochemistry 35*, 11369−11378.
29. Bernstein, F. C., Koetzle, T. F., Williams, G. J., Myer, E. E., Jr., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., and Tasumi, M. (1977) The Protein Data Bank: A computer-based archival file for macromolecular structures, *J. Mol. Biol. 112*, 535−542.
30. Tsai, C. J., and Nussinov, R. (1997) Hydrophobic folding units derived from dissimilar monomer structures and their interactions, *Protein Sci. 6*, 24−42.
31. Tsai, C. J., and Nussinov, R. (1997) Hydrophobic folding units at protein−protein interfaces: Implications to protein folding and protein−protein association, *Protein Sci. 6*, 1426−1437.
32. Pattabiraman, N., Ward, K. B., and Fleming, P. J. (1995) Occluded Molecular Surface: Analysis of Protein Packing, *J. Mol. Recognit. 8*, 334−344.
33. Fleming, P. J., and Richards, F. M. (2000) Protein Packing: Dependence on Protein Size, Secondary Structure and Amino Acid Composition, *J. Mol. Biol. 299*, 487−498.
34. Spiegel, M. R. (1980) *Theory and problems of probability and statistics. SI (metric) edition. Schaum's outline series*, McGraw-Hill Book Co., New Delhi.
35. Kumar, S., Tsai, C. J., and Nussinov, R. (2000) Factors enhancing protein thermostability, *Protein Eng. 13*, 179−191.
36. Makhatadze, G. I., and Privalov, P. (1995) Energetics of protein structure, *Adv. Protein Chem. 47*, 307−425.
37. Myers, J. K., Pace, C. N., and Scholtz, J. M. (1995) Denaturant *m* values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding, *Protein Sci. 4*, 2138−2148.
38. Edgcomb, S. P., and Murphy, K. P. (2000) Structural energetics of protein folding and binding, *Curr. Opin. Biotechnol. 11*, 62−66.
39. Wheelan, S. J., Marchler-Bauer, A., and Bryant, S. H. (2000) Domain size distributions can predict domain boundaries, *Bioinformatics 16*, 613−618.
40. Dengler, U., Siddiqui, A. S., and Barton, G. J. (2001) Protein structural domains: Analysis of the 3Dee domains database, *Proteins 42*, 332−344.
41. Xu, D., and Nussinov, R. (1998) Favorable domain size in proteins, *Folding Des. 3*, 11−17.
42. Alexander, P., Fahnestock, S., Lee, T., Orban, J., and Bryan, P. (1992) Thermodynamic analysis of the folding of the *Streptococcal* protein IgG-binding domains B1 and B2: Why small proteins tend to have high denaturation temperatures, *Biochemistry 31*, 3597−3603.
43. Hiller, R., Zhou, Z. H., Adams, M. W., and Englander, S. W. (1997) Stability and dynamics in a hyperthermophilic protein with melting temperature close to 200 °C, *Proc. Natl. Acad. Sci. U.S.A. 94*, 11329−11332.
44. Rees, D. C., and Adams, M. W. W. (1995) Hyperthermophiles: Taking the heat and loving it, *Structure 3*, 251−254.
45. Beadle, B. M., Baase, W. A., Wilson, D. B., Gilkes, N. R., and Shoichet, B. K. (1999) Comparing the thermodynamic stabilities of a related thermophilic and mesophilic enzyme, *Biochemistry 38*, 2570−2576.
46. Rees, D. C., and Robertson, A. D. (2001) Some thermodynamic implications for the thermostability of proteins, *Protein Sci. 10*, 1187−1194.
47. Kumar, S., Tsai, C. J., and Nussinov, R. (2001) Thermodynamic differences among homologous thermophilic and mesophilic proteins, *Biochemistry 40*, 14152−14165.
48. Shortle, D., and Ackerman, M. S. (2001) Persistence of native-like topology in a denatured protein in 8 M urea, *Science 293*, 487−489.
49. Robertson, A. D., and Murphy, K. P. (1997) Protein structure and energetics of protein stability, *Chem. Rev. 97*, 1251−1267.
50. Tsai, C. J., Maizel, J. V., Jr., and Nussinov, R. (2002) The hydrophobic effect: A new insight from cold denaturation and a two state water structure, *Crit. Rev. Biochem. Mol. Biol. 37*, 55−69.
51. Kumar, S., and Nussinov, R. (1999) Salt bridge stability in monomeric proteins, *J. Mol. Biol. 293*, 1241−1255.
52. Kumar, S., Ma, B., Tsai, C. J., and Nussinov, R. (2000) Electrostatic strengths of salt bridges in thermophilic and mesophilic glutamate dehydrogenase monomers, *Proteins 38*, 368−383.
53. Kannan, N., and Vishveshwara, S. (2000) Aromatic clusters: a determinant of thermal stability of thermophilic proteins, *Protein Eng. 13*, 753−761.
54. Chakravarty, S., and Varadarajan, R. (2002) Elucidation of factors responsible for enhanced thermal stability of proteins: A structural genomics based study, *Biochemistry 41*, 8152−8161.

BI027184+