Thermodynamic Differences among Homologous Thermophilic and Mesophilic Proteins†

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

Laboratory of Experimental and Computational Biology and Intramural Research Support Program- SAIC, Laboratory of Experimental and Computational Biology, National Cancer Institutes 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

ABSTRACT: Here, we analyze the thermodynamic parameters and their correlations in families containing homologous thermophilic and mesophilic proteins which show reversible two-state folding h unfolding transitions between the native and the denatured states. For the proteins in these families, the melting temperatures correlate with the maximal protein stability change (between the native and the denatured states) as well as with the enthalpic and entropic changes at the melting temperature. In contrast, the heat capacity change is uncorrelated with the melting temperature. These and additional results illustrate that higher melting temperatures are largely obtained via an upshift and broadening of the protein stability curves. Both thermophilic and mesophilic proteins are maximally stable around room temperature. However, the maximal stabilities of thermophilic proteins are considerably greater than those of their mesophilic homologues. At the living temperatures of their respective source organisms, homologous thermophilic and mesophilic proteins have similar stabilities. The protein stability at the living temperature of the source organism does not correlate with the living temperature of the protein. We tie thermodynamic observations to microscopics via the hydrophobic effect and a two-state model of the water structure. We conclude that, to achieve higher stability and greater resistance to high and low temperatures, specific interactions, particularly electrostatic, should be engineered into the protein. The effect of these specific interactions is largely reflected in an increased enthalpy change at the melting temperature.

The thermodynamic stability of a protein may vary with the changes in its environmental conditions (e.g., temperature, pH, buffer, salt concentration, presence and absence of chemical denaturants, concentration of the protein, and presence and absence of substrates, ligands, and subunits). The two most common ways of studying protein stability are via thermal and chemical denaturation, using spectroscopy (circular dichroism (CD) and fluorescence) and differential scanning calorimetry (DSC). Thermal-denaturation experiments often yield three thermodynamic parameters: the melting temperature (T_G), the enthalpy change at the melting temperature ($¢H_G$), and the heat capacity change ($¢C_p$) between the native (N) and the denatured (D) states of the protein. These parameters can be used to plot the protein stability curve (*1*, *2*). Protein stability curves describe the temperature-dependent variation of protein stability (the Appendix section gives a description of protein stability curves and the underlying thermodynamic relationships). In studies performed using chemical denaturants such as urea and guanidium hydrochloride (GdnHCl), the Gibbs free energy of unfolding a protein at a given temperature (usually room temperature) is estimated most frequently by using the linear extrapolation method (LEM; *3*). LEM gives the so-called *m* value, the slope of $¢G$ as a function of the concentration of the denaturant measured around the transition state (*3- 5*).

Protein stability curves illustrate that, for all of the proteins that follow a two-state transition, there are two transition

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

† The research of R.N. in Israel has been supported in part by Grant No. 95-00208 from BSF, Israel, by a grant from the Israel Science Foundation administered by the Israel Academy of Sciences, by the Magnet grant, by the Ministry of Science grant, and by the Tel Aviv University Basic Research grants and the Center of Excellence administered by the Israel Academy of Sciences. This project has been funded in whole or in part with Federal funds from the National Cancer Institute, National Institutes of Health, under Contract No. NO1-CO-56000.

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

§ Intramural Research Support Program- SAIC, Laboratory of Experimental and Computational Biology, National Cancer Institute-Frederick.

| Sackler Institute of Molecular Medicine, Tel Aviv University.

[1] Abbreviations: T_G, heat-denaturation (melting) temperature; T'_G, cold-denaturation temperature; T_L, living temperature of the source organism; T_S, temperature of maximal protein stability; $¢H_G$; molar enthalpy change between native and denatured states of a protein at T_G; $¢S_G$, molar entropy change between native and denatured states of a protein at T_G; $¢C_p$, molar heat capacity change between native and denatured states of a protein; $¢G(T)$, molar Gibbs free-energy change between the native and denatured states of a protein at temperature T; $¢G(T_S)$, molar Gibbs free-energy change between the native and denatured states of a protein at T_S; $¢G(T_L)$, molar Gibbs free-energy change between the native and denatured states of a protein at T_L; $¢G$(298 K), molar Gibbs free-energy change between the native and denatured states of a protein at room temperature; EcRnaseH, *Escherichia coli* Ribonuclease H; TtRnaseH, *Thermus thermophilus* Ribonuclease H; °C, degrees Celsius; kcal/mol, kilocalories per mole; Å^2, square angstroms; DSC, differential scanning calorimetry; CD, circular dichroism; r^2, square of the linear correlation coefficient.

10.1021/bi0106383 CCC: $20.00
Published on Web 10/30/2001

temperatures where $\Delta G(T) = 0$. These are T_G and T'_G, termed heat- and cold-denaturation transition temperatures, respectively. Protein heat denaturation is driven by its favorable increase in entropy (*6*), and protein cold denaturation is driven by the favorable decrease in enthalpy (*2*, *7*, *8*). The analyses of both ΔC_p and *m* values have shown that these macroscopic thermodynamic parameters are related to ΔASA, the difference in the accessible surface areas between the denatured and the native protein states (*5*, *9*- *11*).

To *microscopically* understand the *macroscopic* parameters of protein stability, it is useful to consider a two-state model of the water structure. Such a model enables the prediction of cold denaturation, the presence of molten globule (MG) states in heat but not in cold denaturation, and the hydrophobic effect (Tsai, C.-J., Maizel, J. V., and Nussinov, R., unpublished work). The prediction of cold denaturation is in agreement with the prediction by thermodynamically based stability plots and is consistent with the hydrophobic effect and the minimum solubility of small organic solutes around room temperature (*12*- *15*).

Microscopic Scheme. The Gibbs free energy for the cold denatured (D′) $\leftrightarrow$ folded (N) $\leftrightarrow$ heat denatured (D) protein is defined in terms of enthalpy and entropy changes

$$\Delta G = G_D - G_N = \Delta H - T\Delta S$$

For the denatured states to be favorable at the two extreme temperatures ($T \leq T'_G$ and $T \geq T_G$), the entropy and the enthalpy changes in the system (the protein and the surrounding solvent water) should be compensatory. When proteins denature, they preferentially expose their nonpolar surface area. A nonpolar surface has the effect of ordering the first-shell water molecules, reducing the entropy of the system. A two-state water model is useful in understanding the temperature dependence of protein stability. In such a model, water consists of dynamically transforming different intermolecular hydrogen-bonding types (*16*, *17*). The first is the enthalpically favored, "normal" icelike, tetrahedrally connected hexagonal hydrogen-bonding type, with optimal hydrogen-bond networks. The second is the entropically favorable, enthalpically unfavorable, highly fluctuating liquid form. At low temperatures, the hexagonal ice hydrogen-bonding type prevails. At high temperatures, the denser liquid types dominate, with a fluctuating gradient of interconverting hydrogen-bonding types from low to high temperatures. The water structure is dynamic, with the hydrogen bonds continuously broken and created on a very short time scale (*18*).

Below the protein cold-denaturation temperature ($T < T'_G$), the fraction of normal hexagonal ice in solvent water increases and the protein cold denatures because of the loss of the hydrophobic effect. In the denatured state of the protein, the exposure of the nonpolar surface promotes optimal-ordered hydrogen-bond formation at the first shell of the water molecules, propagating to the outer shells. This also explains the clathration cagelike formation of small organic solutes in water. The entropy contribution to the Gibbs free-energy change is negative. However, because the hydrogen-bond networks are optimized, the enthalpy is also reduced, overcoming the unfavorable entropy contribution. Hence, at these temperatures, the denatured state (D′) of the protein is energetically favorable.

At temperatures above the protein heat-denaturation temperature ($T > T_G$), the protein heat denatures because of an increase in the entropy of the system. Liquid water dominates, with favorable entropy and unfavorable enthalpy terms. Order cannot propagate in highly fluctuating water molecules. The increase in the entropy of the system overcomes the enthalpically unfavorable exposure of the nonpolar surface area of the protein, and the denatured state (D) of the protein is energetically favorable. Hence, at temperatures $T \leq T'_G$ or $T \geq T_G$, the significant differences in the relative fractions of normal hexagonal icelike and highly fluctuating denser liquid hydrogen-bonding types in the surrounding water molecules play critical roles in protein denaturation.

Starting from T_G, as the temperature cools, the fraction of hexagonal ice in solvent water gradually increases, enabling a larger extent of order propagation. This would reduce the entropy of the system. Hence, the denatured state of the protein converts into the more-compact molten-globule (MG) state, burying the nonpolar surface area. The MG state has considerable nativelike secondary structure, a compact fold lacking well-defined tertiary interactions, larger solvent surface accessibility, and low cooperativity of thermal unfolding. Those intermediates not conforming to these characteristics are not included in this definition (*19*- *22*). Thus, the D $\rightarrow$ MG step is entropically driven. On the other hand, in the MG $\rightarrow$ N (native-state) folding reaction, the water structure plays an insignificant role. The atomic packing, electrostatic, and disulfide interactions are optimized within the protein. Hence, microscopically, the protein hydrophobic effect is related to the water structure, which is a function of temperature. Macroscopically, it is reflected in ΔASA and is related to ΔC_p. At $T'_G \leq T \leq T_G$, because the protein conformational entropy is reduced in the MG $\rightarrow$ N (*6*) step and the entropy of the water is reduced in the N $\rightarrow$ D′ step, the reduction in the enthalpy of the system should be significant enough to overcome the loss of entropy and drive the reaction to the native state in the first case and to a cold-denatured state in the second. Because MG $\rightarrow$ N is under enthalpy control, the microscopics predict that ΔASA (and ΔC_p) will be uncorrelated with T_G. On the other hand, the formation of specific interactions reflected in ΔH is critically important in heat and cold denaturation.

Macroscopic Analysis: Parameters and Their Correlations. We have collected and analyzed the thermodynamic data relating to families of thermophilic/mesophilic (T/M) proteins. All of these are reversible two-state folding proteins. We present their stability plots and the correlations among their thermodynamic parameters. The melting temperature (T_G), the enthalpy change at the melting temperature (ΔH_G), the entropy change at the melting temperature (ΔS_G), and the maximal protein stability change ($\Delta G(T_S)$) are correlated with one another. These correlations consistently suggest that the higher melting temperatures of thermophilic proteins are obtained by a higher $\Delta G(T_S)$. The increase in $\Delta G(T_S)$ is largely due to the formation of additional specific interactions in the thermophilic proteins as compared to their mesophilic homologues. We find that there is no correlation between ΔC_p and T_G. This observation is particularly interesting because it indicates that an increase in hydrophobicity is not sufficient to improve protein stability. Consistently, we have previously shown that, while different factors contribute to protein thermostability, the higher T_G in thermophiles is best

correlated with an increase in factors that promote specific interactions, such as salt bridges and side chain- side chain hydrogen bonds (*23*). Electrostatic interactions, such as salt bridges and their networks, have favorable electrostatic contributions to the stability of glutamate dehydrogenase from the hyperthermophile *Pyrococcus furiosus* (*24*, *25*).

The proteins in our database show a larger variability in the range of $\Delta G(T_S)$ as compared to $\Delta G(T_L)$, their stabilities at the respective living temperatures of the source organisms. There is no correlation between $\Delta G(T_L)$ and T_L. Specific $\Delta G(T_L)$ values for different proteins may be related to their function. In our dataset, the lowest $\Delta G(T_L)$ is observed for cold-shock proteins and the highest $\Delta G(T_L)$ for archaeal histones, structural proteins responsible for DNA packaging.

MATERIALS AND METHODS

Data Collection and Database Composition. We have performed a literature search using Pubmed to select a database of experimental protein thermodynamic measurements. The search was supplemented by querying the ProTherm database (*26*). Our aim was to collect thermodynamic data on homologous proteins from (hyper)thermophilic and mesophilic organisms. We have collected five such families. These are archaeal histones (*27*), SH3 domain-containing proteins (*28- 31*), cold-shock proteins (*32- 34*), Rnase H (*35*, *36*), and the catalytic domains of cellulases (*37*). These families are termed T/M. The member proteins in the T/M families show high sequence and structural homologies. Together, T/M families contain thermodynamic data on 19 proteins. To compare the results of the analyses on these T/M families, we have also constructed two families of mesophilic proteins. The mesophilic proteins in these two families share sequence or structural homologies to variable extents. Because many of the proteins in these families are only distantly related in sequence, they facilitate the distinguishing between trends related to protein thermostability from phylogenetic differences between thermophiles and mesophiles. We term the mesophilic protein families A/M (all mesophiles). These families consist of five structurally related acylphosphatases (*38- 40*) and the family of Rnase Sa, Sa2, Sa3, Rnase A, Rnase T1, and Barnase (*41*). The selection of the A/M families for this study is arbitrary. The two A/M families contain 11 proteins. For each protein, we have collected three thermodynamic parameters from the literature: ΔH_G, ΔC_p, and T_G.

For most proteins, thermal denaturation involves some irreversibility. The degree of irreversibility has a greater significance for the thermophilic proteins because the thermolabile amino acids in these proteins may undergo covalent modifications at high temperatures. A complete thermal unfolding process for a two-state protein is described more accurately as follows:

$$N \underset{k_2}{\overset{k_1}{\rightleftharpoons}} D \xrightarrow{k_3} I$$

where N, D, and I are the folded native, reversible-denatured, and irreversible-denatured states of the protein, respectively. In thermodynamic experiments, the extent to which protein unfolding follows a two-state mechanism can be measured by the ratio $\Delta H^{cal}/\Delta H^{van'tHoff}$. ΔH^{cal} is the enthalpy change for unfolding determined by DSC. This enthalpy change value is model-independent. $\Delta H^{van'tHoff}$ is determined from thermal-denaturation experiments using CD spectroscopy. Alternatively, it is calculated by (*42*)

$$\Delta H^{van'tHoff} \approx 4RT_G^2 C_{p,max}/\Delta H^{cal}$$

where $C_{p,max}$ is the maximum of the excess heat capacity at T_G and R is the universal gas constant. The calculation of the $\Delta H^{van'tHoff}$ value assumes a two-state folding model. Thus, a value close to unity for $\Delta H^{cal}/\Delta H^{van'tHoff}$ indicates the validity of a two-state folding model for a monomeric protein (*42*, *43*). The presence of isodichroic point(s) in the CD spectra recorded at different temperatures in the transition region also indicates that the protein follows a two-state transition. The reversibility of protein unfolding is often measured by the reproducibility of repetitive DSC (or CD) scans on the same protein sample.

We have accepted the claim of the original experimental publications on the reversible two-state nature of the proteins in our database. We have further noted the $\Delta H^{cal}/\Delta H^{van'tHoff}$ values and the reversibility of protein folding ↔ unfolding transitions wherever available. Further details are given in the original publications.

The accuracy of our analysis depends on the accuracy of the available data. Pace et al. (*41*) have illustrated that the T_G values are accurate to better than ± 1% and that the value for ΔH_G from a van't Hoff analysis can be determined to about ± 5%, with a good agreement between the different laboratories. However, there may be considerable differences in the ΔC_p values. Recently, Pace and co-workers (*41*, *44*) have noted that the reported values of ΔC_p for Rnase A vary from 1.0 to 2.3 kcal/(mol K). Additionally, it is important to note that, while in general ΔC_p is taken to be independent of T, Privalov (*2*, *43*) has shown that ΔC_p decreases at low and high temperatures.

The number of residues in each protein was used to qualitatively estimate the changes in the accessible surface area (ΔASA) between the native and the denatured states of the protein, using the empirical relationship (*9*)

$$\Delta \mathrm{ASA} \approx -907 + 93N_{res}$$

where N_{res} is the number of residues in a protein. Ideally, we should have measured this quantity from the crystal structures of the proteins. However, high-resolution crystal structures are not available for many of the proteins in our dataset. In those cases where these data are available, the atomic coordinates for several residues are missing. In the following, we present a brief description of each protein family.

Archaeal Histones. The study of Li et al. (*27*) describes the unfolding of four recombinant archaeal histones: rHFoB from the mesophile *Methanobacterium formicicum*, rHMfA and rHMfB from the hyperthermophile *Methanothermus fervidus*, and rHPyA1 from *Pyrococcus* strain GB-3a. All four of the histones form dimers with two-state (N_2 ↔ 2U) transitions. The thermodynamic parameters used in this study were reported from thermal-denaturation experiments using DSC and CD spectroscopy. These spectra were recorded at a pH range of 3.0- 5.0 in 0.2 M and 1.0 M KCl. The data used here correspond to a pH of 5.0 and to 0.2 M KCl. The thermal unfolding of these proteins is more than 90%

reversible. The values of the $\Delta H^{cal}/\Delta H^{van'tHoff}$ ratio for rHMfA and rHFoB are 0.23 and 0.33, respectively. These data are unavailable for other recombinant archaeal histones. Ideally, the value of this ratio should have been 0.5 because the histones are dimers. The archaeal histones family is the only family in our database that contains three thermophiles and one mesophile.

SH3 Domain-Containing Proteins. Data on this family were collected from four different studies (*28*- *31*). In total, this family contains data on eight proteins: two (Sac7d and Sso7d) thermophiles and six mesophiles. All eight of the proteins are small single-domain proteins with an SH3 domainlike folding pattern. All of the proteins show two-state folding h unfolding transitions. In the case of Sso7d, CD spectra showed the unfolding to be two-state and mostly reversible (*28*), with the $\Delta H^{cal}/\Delta H^{van'tHoff}$ ratio between 0.92 and 0.95. Sac7d also shows reversible unfolding with two-state behavior ($\Delta H^{cal}/\Delta H^{van'tHoff}$) 0.97) at 0.3 M KCl (*31*). The $\Delta H^{cal}/\Delta H^{van'tHoff}$ ratio for the Btk-SH3 domain is 1.1 (*30*). The thermal unfolding of these proteins has been studied using DSC and CD spectroscopy. The thermodynamic parameters used in this study are from DSC measurements at the neutral pH (7.0). For the mesophilic proteins, the living temperature of the source organisms was taken to be 37 °C.

Cold-Shock Proteins. This family contains data on cold-shock proteins from three organisms: *Escherichia coli* (CspA; *33*), *Bacillus subtilis* (CspB; *34*), and *Thermotoga maritima* (CspTm; *32*). The thermal unfolding of CspTm is 97 (2% reversible, and the $\Delta H^{van'tHoff}/\Delta H^{cal}$ ratio is 1.06 (0.6 (*32*). All of the proteins are monomeric and two-state folders. The thermodynamic parameters were collected from DSC studies at the neutral pH (7.0- 7.5). This family has two mesophilic and one thermophilic proteins.

Rnase H. The family contains thermodynamic data on the cysteine-free mutants of Rnase H from *E. Coli* (EcRnaseH) and *Thermus thermophilus* (TtRnaseH), reported by Hollien and Marqusee (*36*) using CD spectroscopy. The authors have also determined the free energies of unfolding from the GdnHCl denaturation of EcRnaseH and TtRnaseH. They found an excellent agreement between the experimentally determined free-energy values and those calculated using the Gibbs- Helmholtz equation for these proteins. This indicates that a two-state folding model is valid for the cysteine-free variants of Rnase H. For the EcRnaseH cysteine-free mutant, the thermal transition is not reversible in the absence of the denaturant. The thermal unfolding curves of both proteins were recorded at a pH of 5.5 in 5 mM NaAc and 50 mM KCl.

Catalytic Domain of Cellulases. This family contains CD measurement data reported by Beadle et al. (*37*) on the catalytic domains of cellulase E2 from the thermophile *Thermomonospora fusca* ($E2_{cd}$) and the cellulase CenA from the mesophile *Cellulomonas fimi* ($CenA_{P30}$). Both of these domains show reversible two-state folding h unfolding behavior. For both proteins, the buffer is 50 mM KP_i, 225 mM KCl, and 11.25% ethylene glycol at a pH of 6.8.

FiVe Structurally Related Proteins. The family contains thermodynamic data on two highly homologous acylphosphatases [muscle (Muscle Acp) and common type (CT-Acp)] reported by Dobson et al. (*38*, *40*). The data on these proteins are compared with three proteins [EcHpr (*E. coli* Hpr), AdB, and ADA2h] that show structural similarities to muscle Acp (reported in Table 3 of ref *37*). All five of the proteins in the family are mesophilic. The living temperature of the source organisms was taken to be 37 °C. All of the proteins show reversible two-state folding h unfolding transitions.

Rnase Sa, Sa2, Sa3, Barnase, Rnase T1, and Rnase A. Data on these six ribonucleases were taken from Table 1 in ref *41*. All six are mesophilic, and the living temperature of their source organisms was taken to be 37 °C. All of these show reversible two-state folding h unfolding transitions. Rnase A is structurally unrelated to Rnase Sa, Sa2, and Sa3 but has a similar function. Rnase Sa, Sa2, and Sa3 show high sequence and structural homologies among themselves. Barnase and Rnase T1 have low sequence and structural similarities with these proteins. The values of the $\Delta H^{cal}/\Delta H^{van'tHoff}$ ratios are 0.99, 1.00, and 0.96 for Rnase Sa, Sa2, and Sa3, respectively. The thermal denaturation for these proteins is more than 95% reversible (*41*).

The data in the literature are frequently reported in SI and non-SI units. For uniformity, we used the unit for energy as calories. The conversion factor between calories and joules is 1 cal) 4.184 J.

Computation of Linear Correlation Coefficients and t Values. For each pair of thermodynamic parameters (x,y: x_1,y_1, x_2,y_2, ..., x_n,y_n) in our database, we have fitted a line, $y = a + bx$, using the least-squares procedure. The linear correlation coefficient is calculated by

$$r = \frac{(n\sum xy - \sum x \sum y)}{\sqrt{(n\sum x^2 - (\sum x)^2)(n\sum y^2 - (\sum y)^2)}}$$

Our dataset can be regarded as a sample of protein populations. In this sense, the sampling theory of correlation can be used to determine if the correlations observed in our dataset are relevant for proteins in general. We formulate the null hypothesis that the population correlation coefficient (ρ) for a given parameter pair is zero (H_0; $\rho = 0$), while the linear correlation coefficient for the same parameter pair is r in our dataset. The t value is computed to test the null hypothesis by

$$t = r\sqrt{n-2}/\sqrt{1-r^2}$$

where n is the number of proteins in our dataset. For proteins in the T/M families, the null hypothesis is rejected at the 99% level of confidence if $t > 2.60$ (*45*). The rejection of the null hypothesis for a parameter pair indicates that the two parameters are likely to be correlated with each other in proteins.

RESULTS

Thermodynamic Parameters and Protein Stability CurVes for Thermophiles and Mesophiles. Parts a- e of Figure 1 present the stability plots for five homologous T/M protein families. Parts a and b of Figure 2 present the stability plots for two mesophilic (A/M) protein families. Table 1a details the corresponding protein families, their sizes, their ΔASA, and the thermodynamic values collected from the literature. Table 1b lists the temperature of maximal protein stability (T_S), the Gibbs free-energy change for protein unfolding at T_S ($\Delta G(T_S)$), the living temperature of the source organism

Table 1: Structural and Thermodynamic Parameters for Proteins in Our Database and Parameters Derived from Protein Stability Curves

a. Structural and Thermodynamic Parameters for Proteins[a]

protein name	PDB file	N_{res}	¢ASA (Å^2)	¢H_G (kcal/mol)	¢C_p (kcal/(mol K))	T_G (°C)
Archaeal Histones[b]						
rHFoB		134	11555	115.9 (0.1	2.6	74.8 (0.2
rHMfA	1B67	136	11741	164.1 (1.3	2.2	104.1 (0.3
rHMfB		138	11927	150.3 (0.9	1.9	112.8 (0.3
rHPyA1		134	11555	184.4 (3.0	2.4 (0.2	114.1 (0.6
SH3 Domain-Containing Proteins[c]						
Itk		60	4673	42.5	0.8 (0.1	69
Tec		66	5231	40.4	0.7	71
Btk		67	5324	46.8	0.7 (0.1	80
R-Spectrin	1SHG	62	4859	47.1 (2.4	0.8	66 (0.2
Abl	1ABQ	63	4952	46.4 (2.4	0.8(0.1	68.5 (0.2
Fyn		64	5045	55.7 (3.6	0.8 (0.1	70.6 (0.2
Sac7d	1AZP	66	5231	60.2	0.9	90.7
Sso7d	1BF4	64	5045	63.3	0.6	98.5
Cold-Shock Proteins[d]						
CspA	1MJC	70	5603	43.3 (1.7	0.8	57
CspB	1CSP	67	5324	36.8	0.9 (0.2	53.4
CspTm		66	5231	62.6 (3.1	1.1 (0.1	82 (0.2
Cysteine-Free Mutants of Rnase H[e]						
EcRnaseH	2RN2	155	13508	120 (4	2.7 (0.2	66 (1
TtRnaseH	1RIL	166	14531	131 (5	1.8 (0.1	86 (1
Catalytic Domains of Cellulases[f]						
$CenA_{p30}$				107 (3.1	3.8	56.4 (0.3
$E2_{cd}$				190 (14	3.8	72.2 (0.2
Five Structurally Related Proteins[g]						
muscle Acp		98	8207	93.5 (4.7	1.5 (0.2	56.7 (2.0
CT-Acp	2ACY	98	8207	69.3 (3.5	1.5 (0.2	54.0 (2.0
EcHpr	1POH	85	6998	75.8 (1.2	1.5 (0.1	63.6 (0.1
AdB		82	6719	71.5	0.9	74.2
ADA2h		80	6533	47.6	0.9	77
Rnase Sa, Sa2, Sa3, Barnase, Rnase T1, and Rnase A[h]						
Rnase Sa	1RGG	96	8021	97.4 (4.9	1.5 (0.1	48.4 (0.3
Rnase Sa2		97	8114	68.4 (3.4	1.3 (0.1	41.1 (0.3
Rnase Sa3		99	8300	93.6 (4.7	1.6 (0.1	47.2 (0.3
Rnase T1	9RNT	104	8765	105.7	1.7	51.6
Barnase	1RNB	110	9323	126.6	1.8	53.2
Rnase A	1AFK	124	10625	119.4	1.9	62.8

b. Parameters Derived from Protein Stability Curves[i]

protein name	T_S (°C)	¢$G(T_S)$ (kcal/mol)	T_L (°C)	¢$G(T_L)$ (kcal/mol)	¢G(298 K) (kcal/mol)	protein name	T_S (°C)	¢$G(T_S)$ (kcal/mol)	T_L (°C)	¢$G(T_L)$ (kcal/mol)	¢G(298 K) (kcal/mol)
Archaeal Histones											
rHFoB	32.2	7.2	43.0	6.8	7.0	rHMfB	40.2	14.6	83.0	9.4	13.9
rHMfA	35.3	15.5	83.0	7.9	15.1	rHPyA1	44.1	17.2	95.0	8.0	15.8
SH3 Domain-Containing Proteins											
Itk	20.5	3.1	37.0	2.7	3.1	Abl	14.5	3.8	37.0	3.1	3.6
Tec	17.4	3.2	37.0	2.8	3.2	Fyn	6.8	5.3	37.0	4.1	4.9
Btk	22.2	3.9	37.0	3.7	3.9	Sac7d	26.9	5.4	80.0	1.6	5.4
R-Spectrin	12.7	3.8	37.0	3.0	3.6	Sso7d	9.2	8.0	77.0	3.3	7.7
Cold-Shock Proteins											
CspA	5.0	3.5	37.0	2.2	3.0	CspTm	29.4	4.8	80.0	0.3	4.7
CspB	14.3	2.3	37.0	1.5	2.1						
Cysteine-Free Mutants of Rnase H											
EcRnaseH	24.3	7.5	37.0	6.8	7.5	TtRnaseH	20.1	12.4	68.5	5.6	12.4
Catalytic Domains of Cellulases											
$CenA_{p30}$	29.4	4.4	37.0	4.1	4.3	$E2_{cd}$	25.7	13.1	55.0	7.8	13.1
Five Structurally Related Proteins											
muscle Acp	0.9	8.2	37.0	4.7	6.6	AdB	5.2	7.4	37.0	5.8	6.7
CT-Acp	9.8	4.8	37.0	2.9	4.2	ADA2h	25.8	3.6	37.0	3.4	3.6
EcHpr	16.4	5.4	37.0	4.4	5.3						
Rnase Sa, Sa2, Sa3, Barnase, Rnase T1, and Rnase A											
Rnase Sa	- 9.7	9.1	37.0	3.1	5.8	Rnase T1	- 6.5	9.8	37.0	4.2	6.8
Rnase Sa2	- 8.4	5.5	37.0	0.9	3.0	Barnase	- 10.4	12.8	37.0	5.6	8.7
Rnase Sa3	- 7.2	8.2	37.0	2.7	5.3	Rnase A	5.5	10.5	37.0	7.2	9.3

[a] The thermodynamic parameters ¢H_G, ¢C_p, and T_G were obtained from the literature. N_{res} denotes the number of residues in a protein. ¢ASA is the change in accessible surface area between the native and the denatured states of the protein. ¢H_G and ¢C_p are the change in enthalpy at the melting temperature (T_G) and the change in heat capacity between the native and the denatured states, respectively. The availability of X-ray crystal structure data for the proteins in our database is indicated by their protein data bank (PDB) file names. [b] Data from ref *27*. [c] Data from refs *28*- *31*. [d] Data from refs *32*- *34*. [e] Data from refs *35* and *36*. [f] Data from ref *37*. [g] Data from refs *38*- *40*. [h] Data from ref *41*. [i] The thermodynamic parameters, except T_L, were calculated using eqs 1, 6, and 7 in the Appendix section. T_S is the temperature at which the protein is maximally stable, and ¢$G(T_S)$ is the maximal Gibbs free-energy change of unfolding for the proteins. T_L is the living temperature of the source organism, and ¢$G(T_L)$ is the Gibbs free-energy change of unfolding at the living temperature of the source organism. ¢G(298 K) is the Gibbs free-energy change for protein unfolding at 298 K (room temperature).

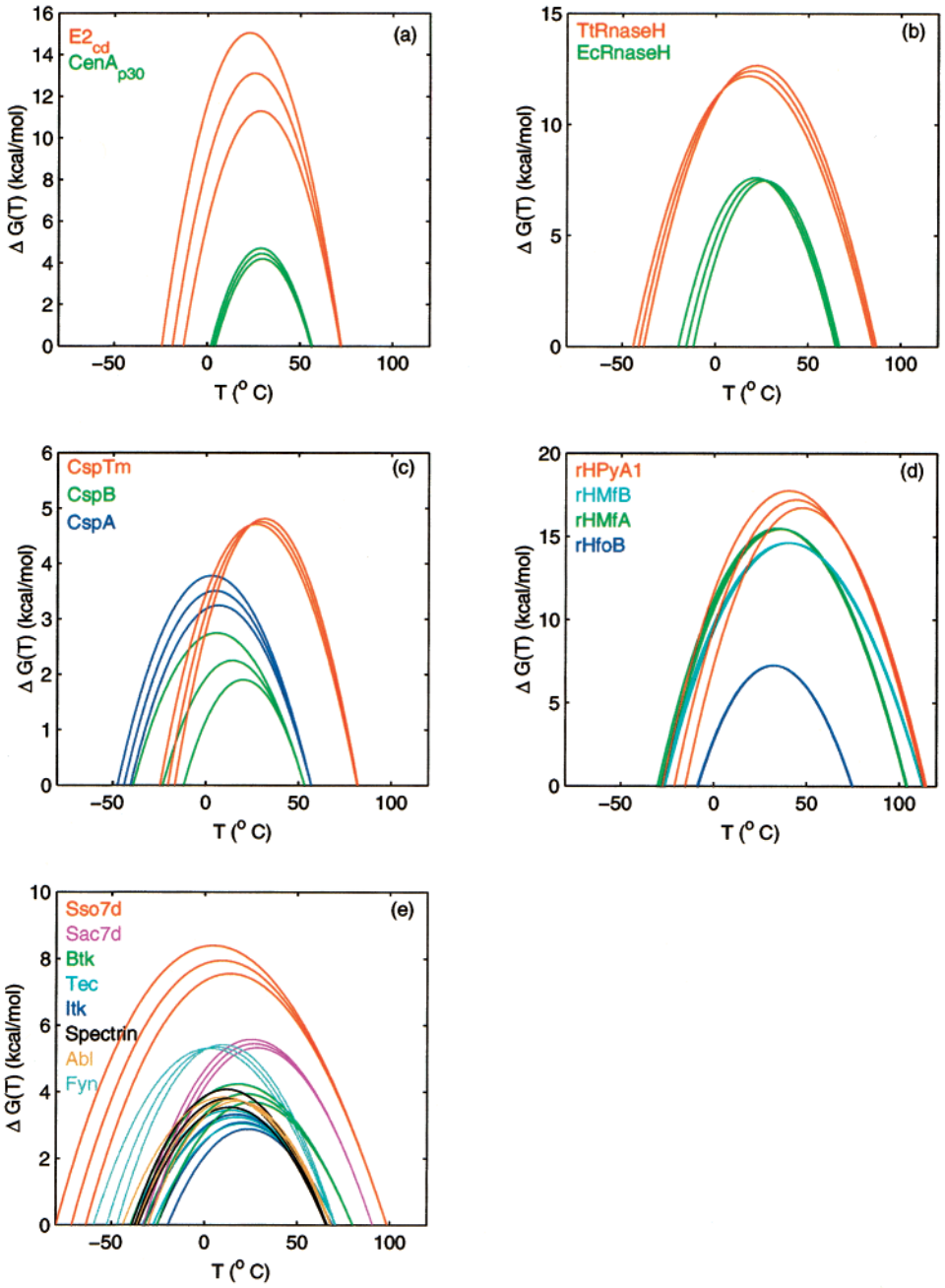

FIGURE 1: Protein stability curves for proteins in T/M (thermophile/mesophile) families. Catalytic domains in (a) cellulases, (b) Rnase H, (c) cold-shock proteins, (d) archaeal histones, and (e) SH3 domain-containing proteins. All of the proteins within a T/M family show a high degree of sequence and structural similarities. In each plot, the *X* axis represents the temperature and the *Y* axis represents the change in Gibbs free energy for protein unfolding ($¢G(T)$). The errors in $¢G(T)$ were estimated using the available experimental errors in $¢H_G$, $¢C_p$, and T_G for each protein.

(T_L), the Gibbs free-energy change for protein unfolding at T_L ($¢G(T_L)$), and the Gibbs free-energy change for protein unfolding at room temperature ($¢G$(298 K)) calculated from the protein stability curves plotted using the Gibbs-Helmholtz equation (see the Appendix section).

In the T/M families, greater protein stability and resistance to higher temperatures are obtained by an upshift and broadening of their protein stability curves as compared to their mesophilic homologues (Figure 1). In most cases, the estimated cold-denaturation temperatures (T'_G) are also lower for the thermophilic proteins. The exception to these observations is the cold-shock protein family, where the protein stability curve for the thermophilic (*T. maritima*) cold-shock protein (CspTm) is both upshifted and right-shifted. These proteins are only marginally stable at the living temperatures of the source organism (Table 1b). For example, $¢G(T_L)$ for CspTm is only 0.35 kcal/mol.

Mesophilic proteins do not show consistent trends (Figure 2). In the A/M family containing ribonucleases, the difference in the maximal stabilities of Barnase and Rnase Sa2 is 7 kcal/mol (Figure 2b and Table 1b). The difference in their melting temperatures is 12.2 °C, and Barnase contains 13 more amino acids than Rnase Sa2. Barnase is only distantly related to Rnase Sa2 (*41*). Hence, the differences in stability for these proteins may be due to phylogeny rather than to thermophilicity.

A higher melting (transition) temperature in a thermophilic protein can be attained in one of three ways (*31*, *37*, *46*,