Protein Structure and the Energetics of Protein Stability

Andrew D. Robertson* and Kenneth P. Murphy*

Department of Biochemistry, The University of Iowa, Iowa City, Iowa 52242

Received March 3, 1997 (Revised Manuscript Received May 14, 1997)

Contents

Andrew D. Robertson was born in Manhattan Beach, CA, in 1959. He received his B.A. in Biology from the University of California at San Diego in 1981 and his Ph.D. in Biochemistry from the University of Wisconsin, Madison, in 1988. After postdoctoral training at Stanford University, he joined the faculty in the Department of Biochemistry at the University of Iowa in 1991, where he is now an Associate Professor. His major research interest is the relationship between protein conformation and the energetics of protein stability and function. Current research is focused on the thermodynamics and kinetics of conformational interconversions in proteins at the level of individual amino acid residues.

Kenneth P. Murphy was born in Lafayette, IN, in 1963. He received his B.A. in Chemistry in 1986 from Metropolitan State College in Denver, CO, and his Ph.D. in Chemistry from the University of Colorado, Boulder, in 1990. Following three years of postdoctoral studies at the Johns Hopkins University, he was appointed Assistant Professor of Biochemistry at the University of Iowa College of Medicine in 1993. His research has focused on understanding the relationship between structure and energetics in protein stability and binding using calorimetry as a primary experimental technique. He was awarded the Stig Sunner Memorial Award by the 50th Calorimetry Conference for his contributions to this field.

I. Introduction

The tendency of proteins to spontaneously adopt a well-defined conformation in solution has intrigued investigators for many decades.[1] The key questions in the study of this *intra*molecular recognition reaction are the same as those driving research into *inter*molecular recognition: what are the molecular determinants of specificity and stability? The distinction between specificity and stability has a long history in studies of intermolecular recognition (e.g., ref 2). In the area of protein folding, this distinction has only recently been articulated in print.[3] In the context of the protein folding reaction, specificity for a given polypeptide chain is reflected in the number of distinct and well-populated conformations adopted by the chain.[4] The majority of native proteins studied to date adopt a specific well-defined conformation. The focus of this review is the relationship between the conformations of such proteins and the energetics of their stability.

The identities of the noncovalent interactions contributing to the stability of the native protein conformation have been established for some time,[5] but considerable debate persists concerning whether and to what extent a given type of interaction favors the native conformation.[6−12] Configurational entropy is widely accepted as the major phenomenon opposing protein stability, but the proposed values of this entropy range from about 17 J K^{-1} mol^{-1} per amino acid residue to about 50 J K^{-1} mol^{-1} per residue.[6,13] In contrast, Honig and Yang propose that the major phenomenon opposing protein stability is desolvation of polar groups upon protein folding.[8] Most researchers agree that the hydrophobic effect plays a key role in stabilizing proteins, but a clear consensus definition of the hydrophobic effect has not been reached.[14−16] Nevertheless, many researchers agree that the hydrophobic effect contributes approximately 8 kJ mol^{-1} per residue, on average, to the free energy

S0009-2665(96)00383-4 CCC: $28.00

of unfolding of proteins at 25 °C.[6,8,17] Hydrogen bonding in proteins has been proposed to be somewhat destabilizing,[8] an indifferent or minor stabilizing force,[11] and a principal contributor to the stability of the native state.[6,9,12,18]

Much of the disagreement derives from the necessity of using models to interpret the thermodynamic data for proteins in terms of specific features of protein structure.[7,9] This follows from the fact that the number of experimental thermodynamic observables in proteins is vanishingly small relative to the thousands of interactions in a typical protein: in the best cases, the thermodynamic data consist of the enthalpy of unfolding (ΔH_u), the entropy of unfolding (ΔS_u), and the heat capacity change upon unfolding (ΔC_p). One can thus deconvolute the energetics of protein stability with respect to atomic-level structure in a number of fundamentally different ways, all of which will be compatible with the primary thermodynamic data.

One approach to increasing and simplifying the information content relative to the thermodynamic data has been to take advantage of the well-documented regularities in native protein structures.[17,19−23] Data for many proteins of known structure have been used to derive empirical relationships between the energetics of protein stability and features of protein structure.[24−27] Similar relationships have been established using thermodynamic data for model compounds, which have served as a basis for interpretation of and comparison with the protein data.[12,28−37]

All approaches to understanding the molecular basis of protein stability ultimately depend on reliable experimental determinations of the thermodynamics of protein unfolding for proteins of known structure. The number of proteins fulfilling this criterion as of late 1996 is more than three times that tabulated by either Privalov and Gill in 1988[38] or Spolar and co-workers in 1992.[27] In seeking relationships between stability and structure, this expanded database presents an opportunity to test the generality of previous observations and the validity of conclusions derived from these observations and, perhaps, to identify trends that were not evident in the smaller collection of proteins.

The focus of this review is on relationships between protein stability and protein structure that can be established with the primary observables, the thermodynamic parameters derived from calorimetric and spectroscopic studies and the structural models derived from X-ray crystallography and NMR spectroscopy. This purely empirical approach will rely on coarse but regular features of structure such as solvent-exposed surface areas, secondary structure content, and numbers of disulfide bonds. The questions at hand are (1) how much information regarding the molecular origins of protein stability can be gleaned from the protein data alone and (2) can these data be used to resolve some of the controversies now in the literature?

II. Determining the Thermodynamics of Unfolding for Globular Proteins

The stability of a globular protein is quantified by the difference in Gibbs energy, ΔG_u, between the denatured state, D, and the native state, N. As the experimental data in this review deal with thermal denaturation, the denatured state is operationally defined as the state of the protein that exists after thermal denaturation. The characteristics of that state, in terms of residual structure, extent of hydration, etc., remain a source of significant speculation and inquiry (see, e.g., refs 39−42).

The equilibrium between the native and denatured states is defined as

$$K = [\mathrm{D}]/[\mathrm{N}] \tag{1}$$

and is related to the ΔG_u as

$$\Delta G_u = -RT \ln K \tag{2}$$

where R is the universal gas constant and T is the absolute temperature. Note that eqs 1 and 2 apply to the equilibrium between the native and denatured states of a protein regardless of the possible presence of intermediate states.

The difference in Gibbs energy is dependent on temperature according to

$$\Delta G_u(T) = \Delta H_u(T) - T\Delta S_u(T) \tag{3}$$

where ΔH_u and ΔS_u are the differences in enthalpy and entropy *at the same temperature at which* ΔG_u *is being evaluated*.

The temperature dependence of ΔH_u and ΔS_u is defined by the heat capacity change, ΔC_p, between the native and denatured states. The change in heat capacity reflects the fact that the amount of heat required to raise the temperature of a solution of unfolded protein is greater than that required for a solution of folded protein of the same concentration. This increase in heat capacity upon unfolding results primarily from restructuring of solvent.[43,44] While ΔC_p is itself slightly temperature dependent,[45] the assumption of a constant ΔC_p does not lead to significant errors in any other parameter.[38] The ΔG_u can thus be described as

$$\begin{aligned}\Delta G_u(T) &= [\Delta H_u(T_R) + \Delta C_p(T - T_R)] - \\ &\qquad T[\Delta S_u(T_R) + \Delta C_p \ln(T/T_R)] \\ &= \Delta H_u(T_R) - T\Delta S_u(T_R) + \\ &\qquad \Delta C_p[(T - T_R) - T \ln(T/T_R)]\end{aligned} \tag{4}$$

where T_R is any convenient reference temperature.

If T_R is equal to T_m, the midpoint for thermal denaturation, then ΔG_u is equal to zero and ΔS_u is just $\Delta H_u/T_m$. Thus eq 4 can be rewritten as

$$\begin{aligned}\Delta G_u(T) = \Delta H_m\left(1 - \frac{T}{T_m}\right) + \Delta C_p[(T - T_m) - \\ T \ln(T/T_m)]\end{aligned} \tag{5}$$

where ΔH_m is the value of ΔH_u at T_m. Equation 5 is generally referred to as the modified Gibbs−Helmholtz equation.

Experimental data are often fit to a modified form of eq 5 in which both sides are divided by $-RT$. Experimental values of ln K as a function of temperature can thus be fitted to yield values for T_m, ΔH_m, and ΔC_p. It must be noted however that such a fit assumes that the experimental values are a true

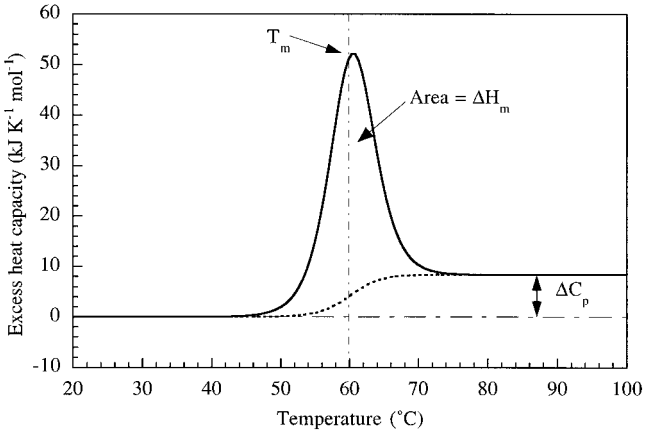

Figure 1. Simulated differential scanning calorimetry experiment for the two-state unfolding of a globular protein. The simulation assumed the following values: T_m = 60 °C, ΔH_m = 418 kJ mol^{-1}, and ΔC_p = 8.4 kJ K^{-1} mol^{-1}.

measure of K, which is only true if there are no stable folding intermediates, as discussed below.

A. Differential Scanning Calorimetry

Differential scanning calorimetry (DSC) is a powerful technique for obtaining data on the thermodynamics of unfolding of globular proteins. Excellent reviews on this technique are available.[46,47] DSC measures the excess heat capacity, $\langle C_p \rangle$, of a protein solution relative to buffer as a function of temperature. The $\langle C_p \rangle$ function can be analyzed to provide the thermodynamic data. As seen in Figure 1, the maximum in $\langle C_p \rangle$ occurs near the T_m of the protein; it occurs directly at T_m only if the ΔC_p of the transition is zero. The area under the $\langle C_p \rangle$ curve gives the ΔH_m of the transition, and the shift in the baseline yields ΔC_p. Thus, in principle, DSC can provide all of the thermodynamics of unfolding for a globular protein in a single experiment.

In practice, it is difficult to obtain good data on the ΔC_p of unfolding from the baseline shift. Instead, several DSC experiments are performed in which the T_m of the protein is perturbed, usually by changing the pH. One then plots ΔH_m as a function of T_m and the slope of this line gives ΔC_p. This analysis assumes that changing pH has no effect on ΔH_u as a function of T; rather the effect of changing pH is assumed to be entirely on ΔS_u.[24] This assumption should be good at low pH because the enthalpies of ionizing acidic groups are generally quite small.[48] The fact that ΔS_u is dependent on pH has important implications for interpreting the unfolding data as discussed below.

One of the most important features of DSC data is that the analysis does not require any assumptions about the presence or absence of stable intermediates in the unfolding process. This is in contrast to optical methods. Consequently, ΔH_m can be readily determined. Additionally, the DSC data can be treated as a progress curve from which one can obtain the van't Hoff enthalpy, ΔH_{vH}, using the same treatment as described for optical methods. Comparison of ΔH_{vH} and the calorimetrically determined ΔH_m can be used to indicate the presence or absence of stable intermediates. The thermodynamic characteristics of such intermediates, if present, can also be deconvoluted from the DSC data.[49]

In spite of the many advantages of studying protein stability by DSC, the technique has several limitations. The sample concentration for typical DSC experiments has needed to be at least 1 mg/mL. With sample volumes of 1−2 mL, this requires that considerable protein be available for study. The high concentrations of protein may lead to difficulties arising from aggregation of the denatured protein or, possibly, self-association of the native state. Accurate DSC studies thus require an assessment of the concentration dependence of the thermodynamics.

Even with moderate concentrations of proteins, it is important to determine that the unfolding transition is reversible before extracting thermodynamic properties from the data. The usual test for reversibility is to perform two DSC scans on each protein and check that the second scan gives all (or most) of the endotherm observed in the first scan. However, the presence of an endotherm upon rescanning the sample is not a test of thermodynamic reversibility, but rather of repeatability. Thermodynamic reversibility requires that the system be at (or very near) equilibrium throughout the reaction. As the equilibrium is being perturbed by scanning in temperature, thermodynamic reversibility in the DSC experiment is better demonstrated by showing that the $\langle C_p \rangle$ function is independent of scan rate. Unfortunately, such tests of reversibility are rarely performed.

In summary, DSC is an excellent method for obtaining thermodynamic data on the unfolding of globular proteins and can provide unique information on the presence and characteristics of stable intermediates. The technique is limited, however, by the requirements for large quantities of protein and high concentrations. Commercial instruments just available within the last year have higher sensitivity and quality data can be obtained from samples at 1/10th the concentration previously required. Such instrumentation will greatly improve the utility of this important technique to protein scientists.

B. Optical Spectroscopy

The thermodynamic parameters for the unfolding of a number of the proteins considered in this review were determined by monitoring thermal and chemical denaturation with spectroscopic techniques.[50−53] For the Arc repressor and HPr, ΔH_m, ΔC_p, and T_m were obtained by the method of Pace and Laurents[54] or that of Chen and Schellman.[55] Both methods rely on detection of cold-induced denaturation or destabilization to obtain estimates for ΔC_p. Thermodynamic parameters for OMTKY3 and iso-1 cyt *c* were obtained in a manner paralleling the usual calorimetric approach: data from individual thermal denaturation experiments were fit to obtain ΔH_m and T_m and variation of pH was used to determine the temperature dependence of ΔH_m, which is described by ΔC_p.

The method of Pace and Laurents entails a combination of chemical and thermal denaturation experiments, with the aim of measuring ΔG_u over a wide range of temperatures.[54] In both thermal and chemical denaturation experiments, ΔG_u is measured over a narrow range of values, ±6 kJ mol^{-1}, where the spectroscopic methods are able to detect changes

in the relative populations of native and denatured protein.[54] The temperature dependence of ΔG_u is then fit to eq 5, the modified Gibbs−Helmholtz equation.

The approach of Chen and Schellman involves thermal denaturation over a sufficient range of temperature to detect heat- and cold-induced denaturation in a single thermal denaturation experiment.[53,55] The data are also fit to the Gibbs−Helmholtz equation (eq 5). In the cases where this approach has been used, chemical denaturants were added in order to observe low- and high-temperature transitions in the same experiment. In principle, the fitted parameters thus reflect the thermodynamics of unfolding only in the presence of denaturant. For HPr, however, the ΔC_p obtained with this approach was identical to that obtained using other procedures.[53] The ΔC_p for the mutant T4 lysozyme studied by Chen and Schellman was 9.1 kJ K^{-1} mol^{-1}, similar to that obtained in the calorimetric study of wild-type protein (Table 1).

One major advantage in the use of spectroscopy over DSC to determine the thermodynamics of protein unfolding is that much less protein is needed in the spectroscopic experiments. Sample concentrations can be as low as 0.01 mg/mL and a wider range of concentrations can be examined, which can serve as a check for self-association reactions. Two significant disadvantages with spectroscopy are the lack of direct measures for intermediates in the unfolding process and the critical role of pre- and posttransition baselines in fitting to obtain the thermodynamic parameters.

The concern about baselines follows from the way in which progress through the unfolding transition is determined: pre- and posttransition baselines are extrapolated into the observable transition zone and the relative concentrations of native and denatured protein are determined from the distances between the observed and extrapolated spectral values.[54] For proper evaluation of fitting errors, terms for baselines should be included in any equation used to fit the spectroscopic data.[56]

Nearly all spectroscopic studies rely on the assumption of a two-state unfolding reaction. Spectroscopic tests for intermediates involve using multiple probes to follow the unfolding reaction,[57] but a negative result is only consistent with, and not proof of, the absence of stable intermediates. It should be noted that issues of repeatability and scan rate dependence discussed above in the context of DSC apply equally to spectroscopic techniques.

C. Precision and Accuracy of Thermodynamic Data

In DSC experiments with modern calorimeters, the least precise variable is probably protein concentration. The sources of uncertainty in determining protein concentration are the precision of a given method, the reproducibility of the method, and systematic deviations between different methods. The results of a recent investigation into various techniques for determining concentrations and extinction coefficients for proteins suggest that, in the best cases, the reproducibility in determining extinction coefficients is about 2%.[58] Thus, the overall experimental precision of the calorimetric data can be no greater than 1 part in 50. In practice, the reproducibility in protein concentration is probably closer to 5%. Previous estimates for the minimum error in determining ΔC_p range from 4% to 10%.[54,59] Reported errors in determining ΔH_m range from 2% to 10%.[60,61]

In principle, the spectroscopic studies of denaturation and van't Hoff analysis of calorimetric data do not depend on knowledge of protein concentration. What is lost in this type of analysis is valuable information concerning the possible presence of stable intermediates. The least precise variable in the spectroscopic studies is likely to be the spectroscopic observable. Although no systematic survey of precision in such measurements has been published, practical experience suggests that, at best, the precision for a given determination may be 1 part in 100; a more accurate value may be 1 part in 20. For both calorimetric and spectroscopic experiments, the overall precision for any determination is probably best assessed by evaluation of the fitting errors.[62]

The question of accuracy in the thermodynamic parameters of unfolding is perhaps best addressed by comparing multiple determinations for the same protein (Tables 1 and 2). To some extent, this will control for some of the systematic errors within laboratories that might be associated with, for example, determining protein concentrations. For nine of the 11 proteins for which there are multiple determinations, experiments have been performed in different laboratories, but usually under similar solution conditions. For the present discussion, relative differences in thermodynamic parameters have been evaluated by dividing the difference between reported values by the smaller of the reported values. Three determinations are available for hen lysozyme and RNase A, so relative differences have been calculated by dividing the standard deviation of the mean by the mean value.

The relative differences in ΔC_p values range from zero to about 80% for whale myoglobin, and the mean relative difference is 14 ± 22%. The relative difference for whale myoglobin is about four times the next largest difference, 19% for RNase A, and the mean relative difference excluding whale myoglobin is 7 ± 6%. This value is very similar to previous estimates for uncertainties in ΔC_p.[54,59] Interestingly, whale myoglobin is the only protein for which the independent determinations have been made under very different solution conditions: one set of experiments were performed at acid pH while the second set were done at alkaline pH.

To facilitate comparison of ΔH_m values obtained at different temperatures, the reported values have been extrapolated to 60 °C and reported as $\Delta H(60)$ in Table 2. While this procedure propagates some of the deviations in ΔC_p values into $\Delta H(60)$, the contributions are generally small because the extrapolations are over a short range of temperature. For the 11 proteins for which multiple determinations have been made, the relative differences in $\Delta H(60)$ values range from 1% for OMTKY3 to 35% for α-lactalbumin. The mean relative difference for multiple determinations is 12 ± 10%, which is in the range of estimated experimental error.[35]

Table 1. Thermodynamics of Unfolding for Globular Proteins of Known Structure

name of protein	pH	T_m, °C	ΔH_m, kJ mol^{-1}	ΔC_p, kJ K^{-1} mol^{-1}	ΔS_m, J K^{-1} mol^{-1}	name of protein	pH	T_m, °C	ΔH_m, kJ mol^{-1}	ΔC_p, kJ K^{-1} mol^{-1}	ΔS_m, J K^{-1} mol^{-1}
α-chymotrypsin[a]	unknown	60	710	12.8	2573	lysozyme (holo equine; transition 2)[bb]	4.5	66.2	133	2.5	393
α-chymotrypsinogen[b]	5	62	619	14.5	1847						
α-lactalbumin[c,d]	5.2	25	−2.5	7.5	−8	lysozyme (hen)[cc]	unknown	60	427	6.3	1281
α-lactalbumin[e,i]	8	25	133	7.6	446	lysozyme (hen)[dd]	unknown	64.05	435	6.4	1289
acyl carrier protein (apo)[f]	6.1	52.7	160	3.3	492	lysozyme (hen)[ee]	2	55	429	6.7	1307
acyl carrier protein (holo)[f]	6.1	64.3	266	6.4	787	lysozyme T4[ff]	2.84	51.2	507	10.1	1562
arabinose binding protein[g]	7.4	59	840	13.2	2528	met repressor[gg]	7	53.2	505	8.9	1547
						myoglobin (horse)[hh]	11.2	62	409	7.6	1220
arc repressor[h,i]	7.3	54	297	6.7	908	myoglobin (whale)[hh]	9.5	85	837	15.6	2336
B1 of protein G[j]	5.4	87.5	258	2.6	715	myoglobin (whale)[ii]	4.75	80.1	575	8.8	1628
B2 of protein G[j]	5.4	79.4	238	2.9	675	OMTKY3[jj]	3.0	72.5	207	2.7	599
barnase[k]	5.5	55.1	500	5.8	1523	OMTKY3[kk]	4.51	85.2	240	2.6	670
barnase[l]	5	53.7	546	6.8	1670	papain[ll]	3.8	83.8	904	13.7	2532
barstar[m]	8	69.9	292	6.2	851	parvalbumin[mm]	7	90	500	5.6	1377
BPTI[n]	4	104	317	2.0	841	pepsin[nn]	5.9	63	1126	18.8	3348
carbonic anhydrase B[o]	unknown	60	725	16.0	2218	pepsinogen[nn]	6	66	1134	24.1	3344
CI2[p]	3.5	73.8	280	2.5	808	plasminogen K4 domain[oo]	7.4	62	315	5.2	940
cyt *b*5 (tryptic fragment)[q]	7	70	332	6.0	968	RNase T1[d,pp]	unknown	25	249	4.9	836
cyt *c* (horse)[r]	unknown	60	393	5.0	1180	RNase T1[qq]	5	61.2	508	4.9	1519
cyt *c* (horse)[s]	unknown	60	307	5.3	922	RNaseA[l]	6	59	372	6.6	1121
cyt *c* (yeast isozyme 1)[t,i]	5	55.4	360	5.7	1096	RNaseA[rr]	5.5	61.9	457	4.8	1365
cyt *c* (yeast isozyme 1)[u]	6	56.2	293	5.2	888	RNaseA[ss]	5.47	64	481	4.8	1360
cyt *c* (yeast isozyme 2)[u]	6	54.5	282	5.2	861	ROP[tt]	6	71	580	10.3	1685
GCN4[v]	7	70	259	3.0	1512	Sac7d[uu]	6	90.9	231	3.6	635
HPr[w,i]	7 (?)	73.4	248	4.9	715	SH3 spectrin[vv]	4	66	197	3.3	581
IL-1β[x]	3	53	351	8.0	1076	*Staphylococcus* nuclease[ww]	7	54	337	9.3	1029
lac repressor headpiece[y]	8	65	118	1.3	349						
lysozyme (human)[z]	4.5	80.3	579	7.2	1638	stefin A[xx]	5	90.8	473	7.4	1300
lysozyme (human)[aa]	2.8	68.8	503	6.6	1470	stefin B[xx]	5	50.2	293	6.7	906
lysozyme (apo equine; transition 1)[bb]	4.5	41.5	154	7.6	488	subtilisin inhibitor[yy]	3.07	50.2	313	8.5	966
						subtilisin BPN′[zz]	8	58.5	370	20.1	1114
lysozyme (apo equine; transition 2)[bb]	4.5	66.44	124	2.6	365	tendamistat[aaa]	∼5	93	307	2.9	838
						thioredoxin[bbb]	7	87.1	411	7.0	1139
lysozyme (holo equine; transition 1)[bb]	4.5	54.73	205	7.4	624	thioredoxin[ccc]	6.5	86.4	444.0	7.4	1235
						trp repressor[ddd]	7.5	90.3	448	6.1	1232
						ubiquitin[eee]	4	90	308	3.3	848

[a] Tischenko, V. M.; Tiktopulo, E. I.; Privalov, P. L. *Biofizika (USSR)* **1974**, *19*, 400. [b] Privalov, P. L.; Khechinashvili, N. N.; Atanasov, B. P. *Biopolymers* **1971**, *10*, 1865. [c] Griko, Y. V.; Freire, E.; Privalov, P. L. *Biochemistry* **1994**, *33*, 1889. [d] The thermodynamics were obtained from a global fit of data and are reported at 25 °C. [e] Xie, D.; Bhakuni, V.; Freire, E. *Biochemistry* **1991**, *30*, 10673. [f] Horvath, L. A.; Sturtevant, J. M.; Prestegard, J. H. *Protein Sci.* **1994**, *3*, 103. [g] Fukada, H.; Sturtevant, J. M.; Quiocho, F. A. *J. Biol. Chem.* **1983**, *258*, 13193. [h] Reference 50. [i] Determined from optically monitored thermal melts. [j] Alexander, P.; Fahnestock, S.; Lee, T.; Orban, J.; Bryan, P. *Biochemistry* **1992**, *31*, 3597. [k] Griko, Y. V.; Makhatadze, G. I.; Privalov, P. L.; Hartley, R. W. *Protein Sci.* **1994**, *3*, 669. [l] Martinez, J. C.; El Harrous, M.; Filimonov, V. V.; Mateo, P. L.; Fersht, A. R. *Biochemistry* **1994**, *33*, 3919. [m] Agashe, V. R.; Udgaonkar, J. B. *Biochemistry* **1995**, *34*, 3286. [n] Makhatadze, G. I.; Kim, K.-S.; Woodward, C.; Privalov, P. L. *Protein Sci.* **1993**, *2*, 2028. [o] Tatunashvili, L. V.; Privalov, P. L. *Biofizika (USSR)* **1986**, *31*, 578. [p] Jackson, S. E.; Moracci, M.; elMasry, N.; Johnson, C. M.; Fersht, A. R. *Biochemistry* **1993**, *32*, 11259. [q] Pfeil, W.; Bendzko, P. *Biochim. Biophys. Acta* **1980**, *626*, 73. [r] Potekhin, S.; Pfeil, W. *Biophys. Chem.* **1989**, *34*, 55. [s] Hagihara, Y.; Tan, Y.; Goto, Y. *J. Mol. Biol.* **1994**, *237*, 336. [t] Reference 52. [u] Liggins, J. R.; Sherman, F.; Mathews, A. J.; Nall, B. T. *Biochemistry* **1994**, *33*, 9209. [v] Thompson, K. S.; Vinson, C. R.; Shuman, J. D.; Freire, E. *Biochemistry* **1993**, *32*, 5491. [w] Reference 53. [x] Makhatadze, G. I.; Clore, G. M.; Gronenborn, A. M.; Privalov, P. L. *Biochemistry* **1994**, *33*, 9327. [y] Hinz, H.-J.; Cossman, M.; Beyreuther, K. *FEBS Letts.* **1981**, *129*, 246. [z] Kuroki, K.; Taniyama, Y.; Seko, C.; Nakamura, H.; Kikuchi, M.; Ikehara, M. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 6903. [aa] Herning, T.; Yutani, K.; Inaka, K.; Kuroki, R.; Matsushima, M.; Kikuchi, M. *Biochemistry* **1992**, *31*, 7077. [bb] Griko, Y. V.; Freire, E.; Privalov, G.; Van Dael, H.; Privalov, P. L. *J. Mol. Biol.* **1995**, *252*, 447. [cc] Cooper, A.; Eyles, S. J.; Radford, S. E.; Dobson, C. M. *J. Mol. Biol.* **1992**, *225*, 939. [dd] Schwarz, F. P. *Thermochim. Acta* **1989**, *147*, 71. [ee] Pfeil, W.; Privalov, P. L. *Biophys. Chem.* **1976**, *4*, 23. [ff] Connelly, P. R.; Ghosaini, L.; Hu, C.-Q.; Kitamura, S.; Tanaka, A.; Sturtevant, J. M. *Biochemistry* **1991**, *30*, 1887. [gg] Johnson, C. M.; Cooper, A.; Stockley, P. G. *Biochemistry* **1992**, *31*, 9717. [hh] Kelly, L.; Holladay, L. A. *Biochemistry* **1990**, *29*, 5062. [ii] Privalov, P. L.; Griko, Y. V.; Venyaminov, S. Y.; Kutyshenko, V. P. *J. Mol. Biol.* **1986**, *190*, 487. [jj] Swint, L.; Robertson, A. D. *Protein Sci.* **1993**, *2*, 2037. [kk] Swint-Kruse, L.; Robertson, A. D. *Biochemistry* **1995**, *34*, 4724. [ll] Tiktopulo, E. I.; Privalov, P. L. *FEBS Lett.* **1978**, *91*, 57. [mm] Filimonov, V. V.; Pfeil, W.; Tsalkova, T. N.; Privalov, P. L. *Biophys. Chem.* **1978**, *8*, 117. [nn] Privalov, P. L.; Mateo, P. L.; Khechinashvili, N. N.; Stepanov, V. M.; Revina, L. P. *J. Mol. Biol.* **1981**, *152*, 445. [oo] Novokhatny, V. V.; Kudinov, S. A.; Privalov, P. L. *J. Mol. Biol.* **1984**, *179*, 215. [pp] Plaza del Pino, I. M.; Pace, C. N.; Freire, E. *Biochemistry* **1992**, *31*, 11196. [qq] Yu, Y.; Makhatadze, G. I.; Pace, C. N.; Privalov, P. L. *Biochemistry* **1994**, *33*, 3312. [rr] Straume, M.; Freire, E. *Anal. Biochem.* **1992**, *203*, 259. [ss] Privalov, P. L.; Tiktopulo, E. I.; Khechinashvili, N. N. *Int. J. Pept. Protein Res.* **1973**, *5*, 229. [tt] Steif, C.; Hinz, H.-J.; Cesareni, G. *Proteins: Struct., Funct., Genet.* **1995**, *23*, 83. [uu] McCrary, B. S.; Edmondson, S. P.; Shriver, J. W. *J. Mol. Biol.* **1996**, *264*, 784. [vv] Viguera, A. R.; Martinez, J. C.; Filimonov, V. V.; Mateo, P. L.; Serrano, L. *Biochemistry* **1994**, *33*, 2142. [ww] Tanaka, A.; Flanagan, J.; Sturtevant, J. M. *Protein Sci.* **1993**, *2*, 567. [xx] Zerovnik, E.; Lohner, K.; Jerala, R.; Laggner, P.; Turk, V. *Eur. J. Biochem.* **1992**, *210*, 217. [yy] Tamura, A.; Kimura, K.; Takahara, H.; Akasaka, K. *Biochemistry* **1991**, *30*, 11307. [zz] Pantoliano, M. W.; Whitlow, M.; Wood, J. F.; Dodd, S. W.; Hardman, K. D.; Rollence, M. L.; Bryan, P. N. *Biochemistry* **1989**, *28*, 7205. [aaa] Renner, M.; Hinz, H.-J.; Scharf, M.; Engels, J. W. *J. Mol. Biol.* **1992**, *223*, 769. [bbb] Santoro, M. M.; Bolen, D. W. *Biochemistry* **1992**, *31*, 4901. [ccc] Ladbury, J. E.; Wynn, R.; Hellinga, H. W.; Sturtevant, J. M. *Biochemistry* **1993**, *32*, 7526. [ddd] Bae, S. J.; Chou, W. Y.; Matthews, K.; Sturtevant, J. M. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 6731. [eee] Wintrode, P. L.; Makhatadze, G. I.; Privalov, P. L. *Proteins: Struct., Funct., Genet.* **1994**, *18*, 246.

The ΔS_m values are dependent upon pH, as described above, and this should introduce an additional source of error when comparing ΔS_m or $\Delta S(60)$ values obtained from independent studies. In

Table 2. Thermodynamic Parameters Used for Regression Analysis[a]

name of protein	ΔC_p	$\Delta H(60)$	$\Delta S(60)$	ΔH^*	ΔS^*
α-chymotrypsin	12.8	709	2570	1230	4420
α-chymotrypsinogen	14.5	590	1760	1180	3860
α-lactalbumin	7.5	260	824	564	1910
α-lactalbumin	7.6	400	1292	708	2400
acyl carrier protein (apo)	3.3	185	566	320	1050
acyl carrier protein (holo)	6.4	238	705	499	1640
arabinose binding protein	13.2	853	2568	1390	4480
arc repressor	6.7	337	1029	608	2000
B1 of protein G	2.6	187	509	292	886
B2 of protein G	2.9	182	511	299	932
barnase	5.8	528	1609	762	2450
barnase	6.8	589	1800	864	2790
barstar	6.2	230	669	483	1570
BPTI	2.0	229	592	310	882
carbonic anhydrase B	16.0	725	2218	1370	4530
CI2	2.5	246	706	347	1070
cyt *b*5 (tryp frag)	6.0	272	790	515	1660
cytochrome *c* (horse)	5.0	393	1180	596	1910
cytochrome *c* (horse)	5.3	307	922	523	1700
cytochrome *c* (yeast iso 1)	5.7	386	1180	617	2000
cytochrome *c* (yeast iso 1)	5.2	312	948	523	1700
cytochrome *c* (yeast iso 2)	5.2	311	947	521	1700
GCN4	3.0	230	668	350	1100
HPr	4.9	183	524	379	1230
IL-1β	8.0	407	1250	731	2410
lac repressor headpiece	1.3	112	330	164	518
lysozyme (human)	7.2	434	1220	724	2250
lysozyme (human)	6.6	444	1300	712	2250
lysozyme (apo equine)[b]	7.6	402	1610	709	2710
lysozyme (holo equine)[b]	7.4	361	1450	661	2530
lysozyme (hen)	6.3	427	1280	682	2190
lysozyme (hen)	6.4	409	1210	668	2140
lysozyme (hen)	6.7	462	1410	733	2380
lysozyme T4	10.1	595	1830	1000	3300
met repressor	8.9	566	1730	928	3030
myoglobin (horse)	7.6	394	1180	703	2280
myoglobin (whale)	15.6	447	1210	1080	3470
myoglobin (whale)	8.8	399	1120	754	2380
OMTKY3	2.7	173	500	283	891
OMTKY3	2.6	175	481	280	857
papain	13.7	578	1590	1130	3570
parvalbumin	5.6	332	894	559	1706
pepsin	18.8	1069	3180	1830	5910
pepsinogen	24.1	989	2910	1970	6410
plasminogen K4 domain	5.2	305	909	516	1670
RNase T1	4.9	419	1380	616	2080
RNase T1	4.9	502	1500	699	2210
RNaseA	6.6	379	1140	645	2090
RNaseA	4.8	462	1300	656	2000
RNaseA	4.8	448	1340	643	2040
ROP	10.3	467	1350	884	2840
Sac7d	3.6	120	316	265	837
SH3 spectrin	3.3	178	523	309	994
Staphylococcus nuclease	9.3	392	1200	767	2540
stefin A	7.4	245	645	545	1720
stefin B	6.7	359	1110	630	2080
subtilisin inhibitor	8.5	395	1220	738	2440
subtilisin BPN′	20.1	400	1210	1214	4120
tendamistat	2.9	212	565	329	985
thioredoxin	7.0	222	596	504	1600
thioredoxin	7.4	249	673	548	1740
trp repressor	6.1	263	701	510	1590
ubiquitin	3.3	208	561	343	1040

[a] For cases in which the proteins are derived from different species, the order here is the same as in Table 1. $\Delta H(60)$ and $\Delta S(60)$ are the ΔH and ΔS of unfolding at 60 °C. ΔH^* is the ΔH of unfolding at 100 °C and ΔS^* is the ΔS of unfolding at 112 °C. All units are as in Table 1. [b] Combined data for transitions 1 and 2.

fact, the range of relative differences in $\Delta S(60)$ values, 4−36%, is similar to that for $\Delta H(60)$. The mean relative difference is 15 (±9)%, which is again quite similar to the mean and standard deviations seen for $\Delta H(60)$. The lack of significant additional uncertainty in $\Delta S(60)$ may result from the fact that most sets of independent determinations were made at similar pH values (Table 1).

III. Correlation of Unfolding Thermodynamics with Protein Structure

A. Database of Unfolding Thermodynamics for Proteins of Known Structure

For this review, the minimal criteria for selection of a protein for consideration are (1) ΔH_m, ΔC_p, and T_m values have been published, (2) the unfolding reaction is reversible, and (3) a structural model for the protein, or a closely related protein, has been deposited in the Protein Data Bank (PDB).[63,64] Thermodynamic parameters for the unfolding of 49 different proteins are assembled in Table 1. For 11 different proteins, at least two independent determinations either from different laboratories or made using alternative methods are included. The ΔH_m and T_m values generally correspond to values obtained under conditions of maximal stability and ΔS_m values have been calculated by dividing ΔH_m by T_m. This database is a work in progress and the authors invite corrections and additions to Table 1.

To put the thermodynamic parameters on a similar footing for correlation with features of protein structure, ΔH_m and ΔS_m at 60 °C ($\Delta H_u(60)$ and $\Delta S_u(60)$) have been calculated using the experimental values and ΔC_p (Table 2). This temperature was chosen because it has been used in previous studies and because it is close to the mean and median T_m values, 65.5 (±2.0) °C and 62.5 °C, respectively, reported in Table 1. Adjustment of ΔH_m and ΔS_m from experimental T_m values to 60 °C means extrapolating over as much as 44 °C, but most experimental T_m values are much closer to 60 °C: the mean deviation of the experimental T_m values from 60 °C is 5.5°.

When seeking patterns in diverse collections of protein structures, two of the most widely used regular features of protein structure are solvent-accessible surface areas[20,25,36,37,65−67] and secondary structure.[19,21,68] Tables 3 and 4 summarize these structural features for the proteins whose thermodynamic parameters are reported in Table 1 and 2. All of the thermodynamic values reported in Table 2 are used in the regression analyses discussed throughout the remainder of the review. In cases where there are multiple thermodynamic entries in Table 2, but a single structural entry in Table 3, each of the experimental entries were regressed against the same structural values. In those cases where multiple structure and thermodynamic entries are given, the thermodynamic entries were regressed against structural entries in the same order in which they are given in Tables 2 and 3.

For the proteins in Table 3, the reported surface area is the sum of the differences (ΔA) between the surface of each residue in the native protein and the solvent accessible surface area of the same type of amino acid residue in an Ala-Xaa-Ala extended tripeptide, corrected for the effects of termini. All carbon atoms are classified as apolar, while all non-carbon atoms are classified as polar. Thus the total change in accessible surface area, ΔA_{tot}, is divided into the change in apolar surface area, ΔA_{ap}, and the change in polar surface area, ΔA_{pol}. For the native

Table 3. Surface Area Changes for the Set of Proteins Used for the Regression Analysis[a]

name of protein	PDB file	N_{res}	ΔA_{ap}, Å^2	ΔA_{pol}, Å^2	ΔA_{tot}, Å^2	name of protein	PDB file	N_{res}	ΔA_{ap}, Å^2	ΔA_{pol}, Å^2	ΔA_{tot}, Å^2
α-chymotrypsin[a]	5CHA	237	13808	8648	22456	met repressor[bb]	1CMB	208	12030	8503	20533
α-chymotrypsinogen[b]	2CGA	245	14012	9127	23139	myoglobin (horse)[cc]	1YMB	153	8884	5523	14407
α-lactalbumin[c]	1HML[d]	123	7027	4719	11746	myoglobin (whale)[dd]	4MBN	153	8873	5927	14800
α-lactalbumin[e]	1ALC[f]	122	6773	4814	11586	myoglobin (whale)	1MBO	153	9143	5679	14822
acyl carrier protein[g]	1ACP	77	3346	2755	6101	OMTKY3[ee]	2OVO	56	2162	1874	4036
arabinose binding protein[h]	1ABE	305	19374	12160	31534	papain[ff]	9PAP	212	13071	8692	21762
arc repressor[i]	1ARR	106	5503	4633	10136	parvalbumin[gg]	5CPV	108	5750	4006	9756
B1 of protein G[j]	1PGB	56	2712	1944	4655	pepsin[hh]	5PEP	326	19584	11717	31301
B2 of protein G[k]	1PGX	56	2981	2117	5098	pepsinogen[ii]	3PSG	365	22811	14298	37108
barnase[l]	1BNI	108	6190	4325	10515	plasminogen K4 domain[jj]	1PMK	78	3801	3408	7209
barnase[l]	1BNJ	109	6137	4281	10417	RNase T1[kk]	9RNT	104	5049	3828	8878
barstar[,m]	1BTA	89	5506	2835	8341	RNase T1[ll]	8RNT	104	5126	3812	8938
BPTI[n]	5PTI	58	2715	1956	4671	RNaseA[mm]	3RN3	124	5802	5468	11271
carbonic anhydrase B[o]	2CAB	256	15949	10591	26540	ROP[nn]	1RPR	126	6195	6737	12932
CI2[p]	1COA	64	3368	2198	5566	Sac7d[oo]	1SAP	66	3357	2509	5866
cyt *b*5 (tryp frag)[q]	1CYO	88	4341	3109	7449	SH3 spectrin[pp]	1SHG	57	3284	1994	5278
cytochrome *c* (horse)[r]	1HRC	104	5716	3788	9504	*Staphylococcus* nuclease[qq]	1STN	136	8049	5173	13222
cytochrome *c* (yeast iso 1)[s]	1YCC	108	5669	4074	9743	stefin A[rr]	1CYV	98	5120	3635	8755
cytochrome *c* (yeast iso 2)[t]	1YEA	112	5630	4320	9950	stefin B[ss]	1STF[tt]	95	5217	3508	8725
GCN4[u]	2ZTA	62	2939	2364	5303	subtilisin inhibitor[uu]	3SIC[vv]	107	4975	3568	8543
HPr[v]	2HPR	87	4555	3035	7590	subtilisin BPN′ [ww]	2ST1	275	15672	10308	25980
IL-1β[w]	6I1B	153	8817	5165	13982	tendamistat[xx]	3AIT	74	3338	2784	6122
lac repressor headpiece[x]	1LCD	51	2291	1622	3913	thioredoxin[yy]	2TRX	108	6317	3464	9781
lysozyme (human)[y]	1LZ1	130	7330	5548	12877	trp repressor[zz]	2WRP	105	6146	4122	10268
lysozyme (hen)	1LYS	129	7024	5315	12339	trp repressor[zz]	3WRP	101	5956	3953	9909
lysozyme (equine)[z]	2EQL	129	7147	5564	12711	ubiquitin[aaa]	1UBQ	76	4112	2606	6717
lysozyme T4[aa]	2LZM	164	9709	6709	16418						

[†] The PDB file identifiers are taken from the Brookhaven Protein Data Bank.[58,59] Number of residues, N_{res}, and ΔA values were determined as described in the text. [a] Blevins, R. A.; Tulinsky, A. *J. Biol. Chem.* **1985**, *20*, 4264. [b] Wang, D.; Bode, W.; Huber, R. *J. Mol. Biol.* **1985**, *185*, 595. [c] Ren, J.; Acharya, K. R.; Stuart, D. I. *J. Biol. Chem.* **1993**, *268*, 19292. [d] X-ray structure is for the human protein. Sequence of the human protein differs from the bovine protein at 31 out of 123 residues. [e] Acharya, K. R.; Ren, J.; Stuart, D. I.; C., P. D.; Fenna, R. E. *J. Mol. Biol.* **1991**, *221*, 571. [f] X-ray structure is for the baboon protein. Sequence of the baboon protein differs from the bovine protein at 37 out of 123 residues. [g] Kim, Y.; Prestegard, J. H. *Proteins: Struct., Func., Genet.* **1990**, *8*, 377. [h] Vyas, N. K.; Quiocho, F. A. *Nature* **1984**, *310*, 381. [i] Bonvin, A. M. J. J.; Vis, H.; Burgering, M. J. M.; Breg, J. N.; Boelens, R.; Kaptein, R. *J. Mol. Biol.* **1994**, *236*, 328. [j] Gallagher, T.; Alexander, P.; Bryan, P.; Gilliland, G. L. *Biochemistry* **1994**, *33*, 4721. [k] Achari, A.; Hale, S. P.; Howard, A. J.; Clore, G. M.; Gronenborn, A. M.; Hardman, K. D.; Whitlow, M. *Biochemistry* **1992**, *31*, 10449. [l] Buckle, A. M.; Henrick, K.; Fersht, A. R. *J. Mol. Biol.* **1993**, *234*, 847. [m] Lubienski, M. J.; Bycroft, M.; Freund, S. M. V.; Fersht, A. R. *Biochemistry* **1994**, *33*, 8866. [n] Wlodawer, A.; Walter, J.; Huber, R.; Sjolin, L. *J. Mol. Biol.* **1984**, *180*, 301. Wlodawer, A.; Nachman, J.; Gilliland, G. L.; Gallagher, W.; Woodward, C. *J. Mol. Biol.* **1987**, *198*, 469. [o] Kannan, K. K.; Ramanadham, M.; Jones, T. A. *Ann. N. Y. Acad. Sci.* **1984**, *429*, 49. [p] Jackson, S. E.; Moracci, M.; elMasry, N.; Johnson, C. M.; Fersht, A. R. *Biochemistry* **1993**, *32*, 11259. [q] Mathews, F. S.; Argos, P.; Levine, M. *Cold Spring Harbor Symp. Quant. Biol.* **1972**, *36*, 387. [r] Bushnell, G. W.; Louie, G. V.; Brayer, G. D. *J. Mol. Biol.* **1990**, *214*, 585. [s] Louie, G. V.; Brayer, G. D. *J. Mol. Biol.* **1990**, *214*, 527. [t] Murphy, M. E. P.; Nall, B. T.; Brayer, G. D. *J. Mol. Biol.* **1992**, *227*, 160. [u] O'Shea, E. K.; Klemm, J. D.; Kim, P. S.; Alber, T. *Science* **1991**, *254*, 539. [v] Liao, D.-I.; Herzberg, O. *Structure* **1994**, *2*, 1203. [w] Clore, G. M.; Wingfield, P. T.; Gronenborn, A. M. *Biochemistry* **1991**, *30*, 2315. [x] Chuprina, V. P.; Rullman, J. A. C.; Lamerichs, R. M. J. N.; Van Boom, J. H.; Boelens, R.; Kaptein, R. *J. Mol. Biol.* **1993**, *234*, 446. [y] Artymiuk, P. J.; Blake, C. C. F. *J. Mol. Biol.* **1981**, *152*, 737. [z] Tsuge, H.; Ago, H.; Noma, M.; Nitta, K.; Sugai, S.; Miyano, M. *J. Biochem.* **1992**, *141*, 111. [aa] Weaver, L. H.; Matthews, B. W. *J. Mol. Biol.* **1987**, *193*, 189. [bb] Rafferty, J. B.; Somers, W. S.; Saint-Girons, I.; Phillips, S. E. V. *Nature* **1989**, *341*, 705. [cc] Evans, S. V.; Brayer, G. D. *J. Mol. Biol.* **1990**, *213*, 885. [dd] Takano, T. In *Methods and Applications in Crystallographic Computing*; Oxford University Press: Oxford, 1984. [ee] Bode, w.; Epp, O.; Huber, R.; Laskowski, M., Jr.; Ardelt, W. *Eur. J. Biochem.* **1985**, *147*, 387. X-ray structure is for silver pheasant which differs from the turkey sequence at one residue. [ff] Kamphuis, I. G.; Kalk, K. H.; Swarte, M. B. A.; Drenth, J. *J. Mol. Biol.* **1984**, *179*, 233. [gg] Swain, A. L.; Kretsinger, R. H.; Amma, E. L. *J. Biol. Chem.* **1989**, *264*, 16620. [hh] Cooper, J. B.; Khan, G.; Taylor, G.; Tickle, I. J.; Blundell, T. L. *J. Mol. Biol.* **1990**, *214*, 199. [ii] Hartsuck, J. A.; Koelsch, G.; Remington, S. J. *Proteins* **1992**, in press. [jj] Padmanabhan, K.; Wu, T.-P.; Ravichandran, K. G.; Tulinsky, A. *Protein Sci.* **1994**, *3*, 898. [kk] Martinez-Oyanedel, J.; Choe, H.-W.; Heinemann, U.; Saenger, W. *J. Mol. Biol.* **1991**, *222*, 335. [ll] Ding, J.; Choe, H.-W.; Granzin, J.; Saenger, W. *Acta Crystallogr., Sect. B* **1992**, *48*, 185. [mm] Howlin, B.; Moss, D. S.; Harris, G. W. *Acta Crystallogr., Sect. A* **1989**, *45*, 851. [nn] Eberle, W.; Pastore, A.; Sander, C.; Roesch, P. *J. Biomol. NMR* **1991**, *1*, 71. [oo] Edmondson, S. P.; Qiu, L.; Shriver, J. W. *Biochemistry* **1995**, *34*, 13289. [pp] Musacchio, A.; Noble, M.; Pauptit, R.; Wierenga, R.; Saraste, M. *Nature* **1992**, *359*, 851. [qq] Hynes, T. R.; Fox, R. O. *Proteins: Struct., Funct., Genet.* **1991**, *10*, 92. [rr] Tate, S.; Ushioda, T.; Utsunomiya-Tate, N.; Shibuya, Y.; Ohyama, Y.; Nakano, Y.; Kaji, H.; Inagaki, F.; Samejima, T.; Kainosho, M. *Biochemistry* **1995**, *34*, 14637. [ss] Stubbs, M. T.; Laber, B.; Bode, W.; Huber, R.; Jerala, R.; Lenarcic, B.; Turk, V. *EMBO J.* **1990**, *9*, 1939. [tt] Taken from the complex with papain. [uu] Takeuchi, Y.; Noguchi, S.; Satow, Y.; Kojima, S.; Kumagai, I.; Miura, K.-I.; Nakamura, K. T.; Mitsui, Y. *Protein Eng.* **1991**, *4*, 501. [vv] Taken from the complex with subtilisin. [ww] Bott, R.; Ultsch, M.; Kossiakoff, A.; Graycar, T.; Katz, B.; Power, S. *J. Biol. Chem.* **1988**, *263*, 7895. [xx] Billeter, M.; Schaumann, T.; Braun, W.; Wüthrich, K. *Biopolymers* **1990**, *29*, 695. [yy] Katti, S. K.; LeMaster, D. M.; Eklund, H. *J. Mol. Biol.* **1990**, *212*, 167. [zz] Lawson, C. L.; Zhang, R.-G.; Schevitz, R. W.; Otwinowski, Z.; Joachimiak, A.; Sigler, P. B. *Proteins: Struct., Funct., Genet.* **1988**, *3*, 18. [aaa] Vijay-Kumar, S.; Bugg, C. E.; Cook, W. J. *J. Mol. Biol.* **1987**, *194*, 531.

structure, the algorithm of Lee and Richards,[65] as implemented in the program ACCESS (Scott R. Presnell, University of California at San Francisco), has been used to determine the solvent-accessible surface area using a probe radius of 1.4 Å and a slice width of 0.25 Å. The calculations use whole-atom atomic radii, i.e., hydrogen atoms are not considered explicitly but instead are included by using slightly increased atomic radii for atoms covalently bonded to hydrogens.[65] Consequently, hydrogens from NMR-derived structures are ignored in the calculation.

The appropriate solvent-accessible surface area for the denatured protein is a subject of continuing discussion.[69] The use of a single standard model for

Table 4. Seconday Structure and Disulfide Bonds in the Set of Globular Proteins[a]

name of protein	no. of disulfides	total helix, %	strand, %	turn, %	other, %
α-chymotrypsin	5	11.9	33.5	36.9	17.8
α-chymotrypsinogen	5	11.4	33.5	28.6	26.5
α-lactalbumin	4	47.2	8.9	21.1	22.8
α-lactalbumin	4	38.5	6.6	25.4	29.5
acyl carrier protein	0	26.0	0.0	42.9	31.2
arabinose binding protein	0	45.6	20.7	12.1	21.6
arc repressor	0	26.4	4.7	11.3	57.5
B1 of protein G	0	26.8	42.9	14.3	16.1
B2 of protein G	0	26.8	46.4	14.3	12.5
barnase	0	24.1	24.1	25.0	26.9
barnase	0	23.9	22.9	27.5	25.7
barstar	0	47.2	18.0	18.0	16.9
BPTI	3	20.7	24.1	13.8	41.4
carbonic anhydrase B	0	16.4	31.3	25.8	26.6
CI2	0	17.2	28.1	35.9	18.8
cyt *b*5 (tryp frag)	0	35.2	21.6	26.1	17.0
cytochrome *c* (horse)	0	35.6	3.8	28.8	31.7
cytochrome *c* (yeast iso 1)	0	34.3	3.7	26.9	35.2
cytochrome *c* (yeast iso 2)	0	39.3	3.6	22.3	34.8
GCN4	0	93.5	0.0	0.0	6.5
HPr	0	39.1	27.6	16.1	17.2
IL-1β	0	5.2	47.1	30.1	17.6
lac repressor headpiece	0	56.9	0.0	11.8	31.4
lysozyme (human)	4	43.8	10.8	33.8	11.5
lysozyme (hen)	4	41.1	9.3	35.7	14.0
lysozyme (equine)	4	43.4	9.3	31.0	16.3
lysozyme T4	0	66.5	8.5	5.5	19.5
met repressor	0	23.1	6.3	4.3	66.3
myoglobin (horse)	0	79.1	0.0	9.8	11.1
myoglobin (whale)	0	80.4	0.0	8.5	11.1
myoglobin (whale)	0	83.8	0.0	8.5	7.7
OMTKY3	3	19.6	17.9	25.0	37.5
papain	3	31.1	17.9	12.3	38.7
parvalbumin	0	54.6	0.0	25.9	19.4
pepsin	3	13.2	42.3	23.0	21.5
pepsinogen	3	20.8	38.4	21.4	19.5
K4 frag plasminogen	3	0.0	14.1	50.0	35.9
RNase T1	2	16.3	27.9	29.8	26.0
RNase T1	2	16.3	27.9	29.8	26.0
RNaseA	4	22.6	33.1	23.4	21.0
ROP	0	40.5	0.0	4.0	55.6
Sac7d	0	30.3	40.9	10.6	18.2
SH3 spectrin	0	5.3	47.4	19.3	28.1
Staphylococcus nuclease	0	29.4	30.1	20.6	19.9
stefin A	0	7.1	33.7	37.8	21.4
stefin B	0	22.1	38.9	14.7	24.2
subtilisin inhibitor	2	15.9	33.6	34.6	15.9
subtilisin BPN′	0	30.2	17.1	28.4	24.4
tendamistat	2	0.0	45.9	25.7	28.4
thioredoxin	1	35.2	26.9	24.1	13.9
trp repressor	0	80.0	0.0	0.0	20.0
trp repressor	0	83.2	0.0	5.9	10.9
ubiquitin	0	25.0	30.3	23.7	21.1

[a] The number of residues in a secondary structure class was calculated using STRIDE and converted to percentages. The total helix percentage includes both α and 3_{10} helices. The PDB files and references are in the same order as listed in Table 3.

the denatured state will control for *systematic* errors in the use of a model for the denatured state,[69] but it will not account for any real differences in the extent to which the denatured forms of different proteins may vary in their relative solvent accessibilities.

The assignment of secondary structure in proteins is somewhat dependent on the choice of algorithm.[70,71] Secondary structure is generally defined as a regularly repeating conformation of the polypeptide chain. All algorithms yield very similar results for any given protein but specific criteria for identifying regularities in polypeptide conformation vary from laboratory to laboratory.[70,71] All secondary structure contents reported in Table 4 were assessed using the STRIDE algorithm of Frishman and Argos;[71] use of a single algorithm minimizes variations resulting from the different criteria and algorithms used to derive secondary structure as reported in the PDB files.

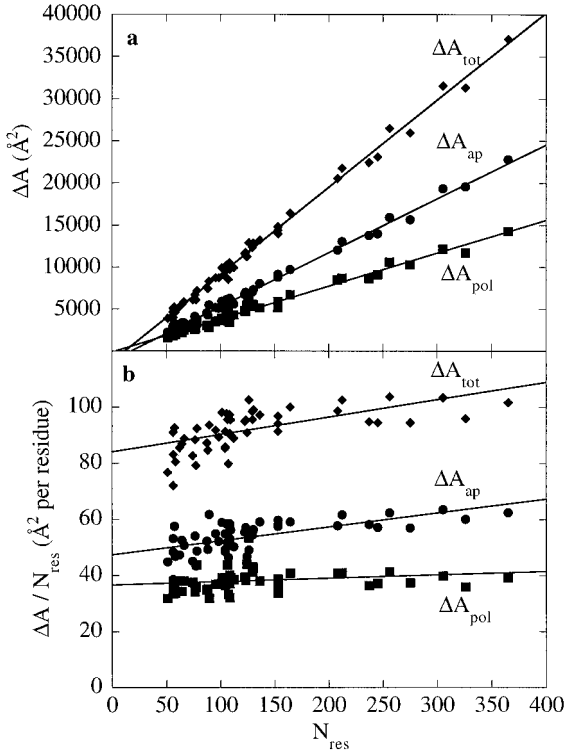

Figure 2. Correlation of surface area changes with protein size. (a) The total change in accessible surface area, ΔA_{tot}, as well as the apolar, ΔA_{ap}, and polar, ΔA_{pol}, contributions are plotted vs the number of residues, N_{res}. The lines are the linear regressions. The slope, intercept, and R^2 values are 104, −1200, and 0.993 for ΔA_{tot}, 64, −1120, and 0.989 for ΔA_{ap}, and 39, −84, and 0.971 for ΔA_{pol}. (b) The change in accessible surface area per residue is plotted vs the number of residues. The lines are the linear regressions. The slope, intercept, and R^2 values are 0.062, 84, and 0.376 for ΔA_{tot}, 0.050, 47, and 0.379 for ΔA_{ap}, and 0.012, 37, and 0.045 for ΔA_{pol}.

The STRIDE algorithm was designed to more closely mimic the secondary structure assignments reported by investigators in the PDB files than the commonly used DSSP algorithm of Kabsch and Sander.[72] In general, the two algorithms yield very similar results: a survey of 226 proteins shows the highest level of disagreement for an individual protein was 14% of the residues.[71]

B. Relationships between Unfolding Thermodynamics and Features of Protein Structure

1. General Structural Features

Before looking at correlations between energetic and structural features of proteins, it is worth examining the correlations of the structural features themselves. For example, it has long been known that the buried surface area correlates with the size of the protein.[20] This is illustrated in Figure 2a in which ΔA_{tot}, ΔA_{ap}, and ΔA_{pol} are plotted vs the number of residues in the protein. In addition to the increase in the total surface area buried in the

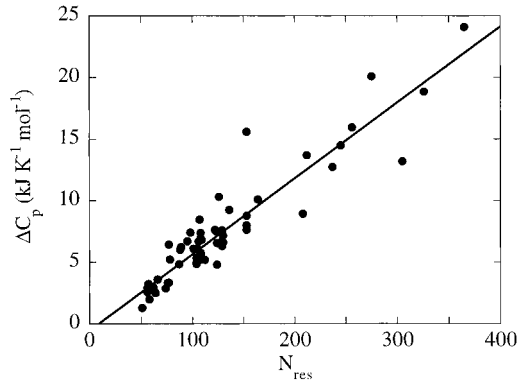

Figure 3. Correlation of ΔC_p of unfolding with the number of residues. The line is the linear regression with slope, intercept, and R^2 of 0.062, −0.53, and 0.862.

Table 5. Results of Regression Analysis of Thermodynamics of Protein Unfolding

thermodynamic parameter	regression variables	regressed values	R^2
ΔC_p	N_{res}	58 ± 1 J K^{-1} (mol res)$^{-1}$	0.859
ΔC_p	ΔA_{tot}	0.61 ± 0.02 J K^{-1} (mol Å^2)$^{-1}$	0.856
ΔC_p	ΔA_{ap}	0.66 ± 0.21 J K^{-1} (mol Å^2)$^{-1}$	0.856
	ΔA_{pol}	0.52 ± 0.32 J K^{-1} (mol Å^2)$^{-1}$	
ΔH (60 °C)	N_{res}	2.92 ± 0.08 kJ (mol res)$^{-1}$	0.766
ΔH (60 °C)	ΔA_{tot}	30.2 ± 0.9 J (mol Å^2)$^{-1}$	0.735
ΔH (60 °C)	ΔA_{ap}	−8 ± 11 J (mol Å^2)$^{-1}$	0.775
	ΔA_{pol}	86 ± 17 J (mol Å^2)$^{-1}$	
ΔH (100 °C)	N_{res}	5.28 ± 0.09 kJ (mol res)$^{-1}$	0.918
$\Delta S°$ (60 °C)	N_{res}	8.8 ± 0.3 J K^{-1} (mol res)$^{-1}$	0.744
$\Delta S°$ (60 °C)	ΔA_{tot}	0.091 ± 0.003 J K^{-1} (mol Å^2)$^{-1}$	0.716
$\Delta S°$ (60 °C)	ΔA_{ap}	−0.03 ± 0.04 J K^{-1} (mol Å^2)$^{-1}$	0.757
	ΔA_{pol}	0.27 ± 0.06 J K^{-1} (mol Å^2)$^{-1}$	
$\Delta S°$ (60 °C)	N_{res}	9.2 ± 4.6 J K^{-1} (mol res)$^{-1}$	0.771
	ΔA_{ap}	−0.11 ± 0.05 J K^{-1} (mol Å^2)$^{-1}$	
	ΔA_{pol}	0.15 ± 0.08 J K^{-1} (mol Å^2)$^{-1}$	
$\Delta S°$ (112 °C)	N_{res}	17.3 ± 0.3 J K^{-1} (mol res)$^{-1}$	0.919

protein with increasing protein size, the surface area buried per residue also increases.[17] As noted previously,[20,73] and as seen in Figure 2b, this increase is mainly due to an increase in the apolar surface buried per residue, while the polar surface buried per residue remains fairly constant. Because the polar surface area buried per residue is nearly constant with protein size while the apolar area per residue increases, the fraction of the buried surface area that is apolar increases with size, but this trend is very weak. Ignoring this weak upward trend, the percentage of the total buried area that is apolar is 58.3 ± 3.4%. The percentage of the total buried area that is polar is 41.7 ± 3.4%.

2. Heat Capacity of Unfolding

The thermodynamic term which has been most often scaled to structural features is ΔC_p.[26,32,35,74−78] The simplest approach is to assume that ΔC_p scales only with the size of the protein, that is with the number of amino acid residues, N_{res}. Regression of ΔC_p on N_{res} (Figure 3) gives a value of 58 ± 2 J K^{-1} (mol res)$^{-1}$ with $R^2 = 0.859$ (Table 5).

The next simplest assumption is that ΔC_p scales with the total change in ASA, ΔA_{tot}. This is nearly the same as regression on N_{res} since N_{res} and ΔA_{tot} are highly correlated ($\Delta A_{tot} = (96.2 \pm 0.7)N_{res}$; $R^2 = 0.987$; Table 5). Regression of ΔC_p on ΔA_{tot} gives 0.61 ± 0.02 J K^{-1} (mol Å^2)$^{-1}$ with $R^2 = 0.856$. Results of model compound studies clearly demonstrate that polar and apolar ASA make different contributions to ΔC_p.[27,30,79] However, simultaneous regression of ΔC_p on both ΔA_{ap} and ΔA_{pol} yields very similar values, 0.66 ± 0.21 and 0.52 ± 0.32 J K^{-1} (mol Å^2)$^{-1}$ respectively, with $R^2 = 0.856$. Thus, as noted by Myers et al.,[77] the separate contributions of polar and apolar surface to ΔC_p are not evident in the protein data.

As noted above, the correlation between structural features and ΔC_p has been investigated previously. The values observed here for 49 different proteins represent a much larger data set than has been used previously. The analysis of Spolar et al.[27] used the set of 12 proteins tabulated by Privalov and Gill,[38] while the more recent analysis by Myers et al.[77] used a set of 26 proteins.

Myers et al. also found that ΔC_p correlated equally well with N_{res} as with ΔA_{tot} or ΔA_{ap} and ΔA_{pol}. Their value per residue was 59.4 J K^{-1} (mol res)$^{-1}$, similar to the 58 J K^{-1} (mol res)$^{-1}$ found here and the 59 J K^{-1} (mol res)$^{-1}$ found by Privalov and Gill.[38] The correlation of ΔC_p with ΔA_{tot} observed by Myers et al. gave a value of 0.79 J K^{-1} (mol Å^2)$^{-1}$ compared to our 0.61 J K^{-1} (mol Å^2)$^{-1}$. The discrepancy between these values probably reflects the different algorithms used in calculating ASA.

The analysis here shows no significant difference in the contribution of apolar and polar surface to ΔC_p, although they have been observed previously to be of opposite sign.[27,30,79] Murphy and Freire[35] found a value of 1.9 J K^{-1} (mol Å^2)$^{-1}$ for apolar surface and −1.1 J K^{-1} (mol Å^2)$^{-1}$ for polar surface based on data for the dissolution of cyclic dipeptides.[79] Spolar et al.[27] found a value of 1.4 J K^{-1} (mol Å^2)$^{-1}$ for apolar surface and −0.67 J K^{-1} (mol Å^2)$^{-1}$ for polar surface from analysis of unfolding data on a set of 14 globular proteins. Finally, Myers et al.[77] find a value of 1.2 J K^{-1} (mol Å^2)$^{-1}$ for apolar surface and −0.38 J K^{-1} (mol Å^2)$^{-1}$ for polar surface from their data set of 26 proteins.

In the data set presented here, the surface area buried by the average protein is 58.3% apolar and 41.7% polar. If the average contribution to ΔC_p per unit surface area is calculated from the apolar and polar contributions weighted by these percentages, all of the above treatments give similar values. The coefficients of Murphy and Freire give a weighted average of 0.65 J K^{-1} (mol Å^2)$^{-1}$; those of Spolar et al. give 0.54 J K^{-1} (mol Å^2)$^{-1}$; those of Myers et al. give 0.54 J K^{-1} (mol Å^2)$^{-1}$; and those from this study give 0.60 J K^{-1} (mol Å^2)$^{-1}$. These can be compared to the value of 0.61 J K^{-1} (mol Å^2)$^{-1}$ from the regression on ΔA_{tot}. This observation points to the difficulty of obtaining coefficients for the apolar and polar contributions to ΔC_p from the protein data alone. Any number of combinations of coefficients can give equally good fits to the data. This is not in conflict with the model compound data, which clearly indicate different contributions of apolar and polar surface to ΔC_p, but merely points to the limitations of using the protein data by themselves.

3. Convergence Temperatures

In the mid-1970s Privalov and co-workers noted an interesting feature in the thermodynamics of unfolding of globular proteins; namely, the ΔH_u values,

when normalized to the molecular weight of the proteins (or to the number of residues), converge to a common value at some high temperature (designated T_H*).[80] Likewise, the normalized ΔS_u values converge to a common value at some high temperature (T_S*).[80] It was originally surmised that the hydrophobic contributions to ΔH_u and ΔS_u approach zero at their respective "convergence temperatures",[80] so that the convergence behavior could be used to determine the contributions of fundamental interactions to the stability of globular proteins. Thus the value of ΔH_u at T_H*(ΔH*) could be attributed to polar and van der Waals interactions, while the value of ΔS_u at T_S*(ΔS*) could be attributed primarily to configurational entropy. The convergence behavior has been the source of significant interest and speculation since that time.[13,27,33,34,38,81−85]

In 1986, Baldwin noted that T_S* for proteins occurs at the same temperature at which the $\Delta S°$ of dissolution for hydrophobic liquids extrapolates to zero[13] suggesting that at T_S* the hydrophobic contribution to ΔS_u was zero. Subsequently, it was shown that the convergence behavior could be analyzed by plotting the normalized ΔS_u (or ΔH_u) vs the normalized ΔC_p.[33] Consider the standard equation for ΔS_u as a function of temperature:

$$\Delta S_u = \Delta S^* + \Delta C_p \ln(T/T_S^*) \quad (6)$$

For a given protein, eq 6 describes the ΔS_u as a function of temperature. However, for a set of proteins a plot of ΔS_u vs ΔC_p will have a slope of ln (T/T_S*) and an intercept of ΔS*.[33]

Similarly, for ΔH_u we have

$$\Delta H_u = \Delta H^* + \Delta C_p(T - T_H^*) \quad (7)$$

Again, eq 7 describes the temperature dependence of ΔH_u for a single protein. A plot of ΔH_u vs ΔC_p for a set of proteins which show convergence will have a slope of ($T - T_H$*) and an intercept of ΔH*.

Using these plots, T_S* was found to be the same for transfer of hydrophobic compounds from the gas, liquid, and solid phases, as well as for protein unfolding.[33] This confirmed the observation of Baldwin that the hydrophobic contribution to ΔS_u approached zero at T_S*. By analogy, it was argued that the hydrophobic contribution to ΔH_u also approached zero at T_H*.[33]

Convergence behavior has also been seen for the aqueous dissolution of a homologous series of model compounds.[79,85] From these, the requirements for observing convergence behavior also have been clarified.[34,79,82] Convergence will be observed for a set of compounds if the following conditions hold: (1) the series is homologous, i.e., one functional group is constant throughout the series of compounds while another functional group varies (e.g., the normal alcohols); and (2) the contributions of the functional groups are independent and additive. If these two conditions are met, then both ΔH and ΔS for the entire set will converge to common values at some temperature. Furthermore, if the series is variable in the number of methylene groups, then convergence occurs at the temperature where the methylene contributions (i.e., the hydrophobic contribution) to the enthalpy or entropy are zero. The convergence

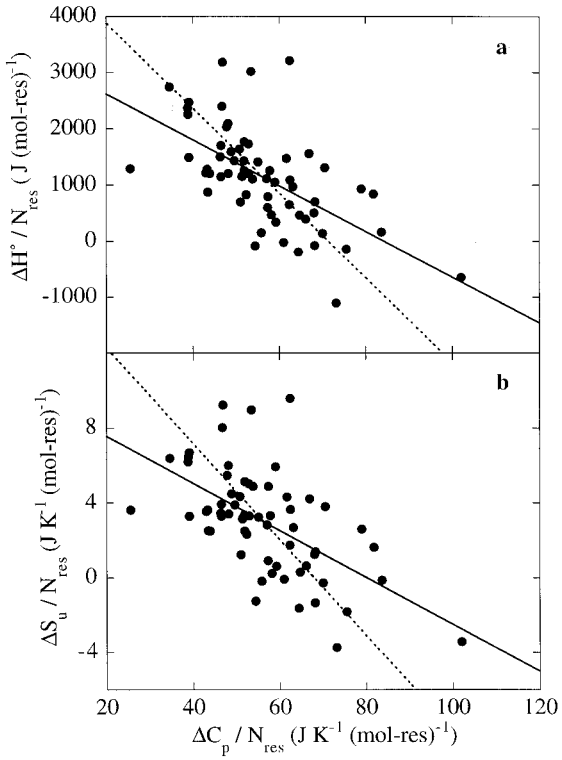

Figure 4. Correlation of the residue normalized ΔH_u (a) and ΔS_u (b) with ΔC_p at 25 °C. The solid lines are the linear regression. The slopes are related to the convergence temperatures, T_H* and T_S*, and the intercepts give the convergence values ΔH* and ΔS*. For ΔH_u, the slope, intercept, and R^2 values are −40.9, 3440, and 0.362, corresponding to a T_H* value of 65.9 °C. The dotted line assumes the previously determined value of T_H* of 100.5 °C. For ΔS_u the slope, intercept, and R^2 values are −0.126, 10.1, and 0.330 corresponding to a T_S* value of 65.0 °C. The dotted line assumes the previously determined value of T_S* of 112 °C.

value, ΔH* or ΔS*, is the contribution of the invariant group to ΔH or ΔS at the convergence temperature.[79,85] Consequently, if convergence is observed, the thermodynamics can be parsed into two groups, those arising from apolar interactions and those arising from other contributions.[34,79]

The globular proteins appear to represent a homologous series of compounds when the surface areas are normalized per residue, as indicated in Figure 2b above. The polar surface area buried per residue is essentially constant with increasing protein size, whereas the apolar surface area buried per residue increases. However, the average value of ΔA_{pol} per residue is 38.3 ± 3.9 Å^2 res^{-1} while the average value of ΔA_{ap} per residue is 53.6 ± 5.5 Å^2 res^{-1}. Looking at the standard deviations, the variability in buried apolar area is only slightly greater than that for polar surface area.

To determine if the proteins in Table 1 exhibit convergence behavior, plots of ΔH_u and ΔS_u at 25 °C vs ΔC_p, all normalized per residue, were constructed (Figure 4). The solid line represents the linear regression of the data. The linear regression of the ΔH_u data gives a slope of −40.9 and an intercept of 3440 with $R^2 = 0.36$. This corresponds to T_H* equal to 65.9 °C and ΔH* equal to 3.44 kJ (mol res)$^{-1}$. The dotted line corresponds to the analysis of a smaller data set by Murphy and Gill[34] with T_H* = 100.5 °C and ΔH* = 5.64 kJ (mol res)$^{-1}$. The set of Spolar et

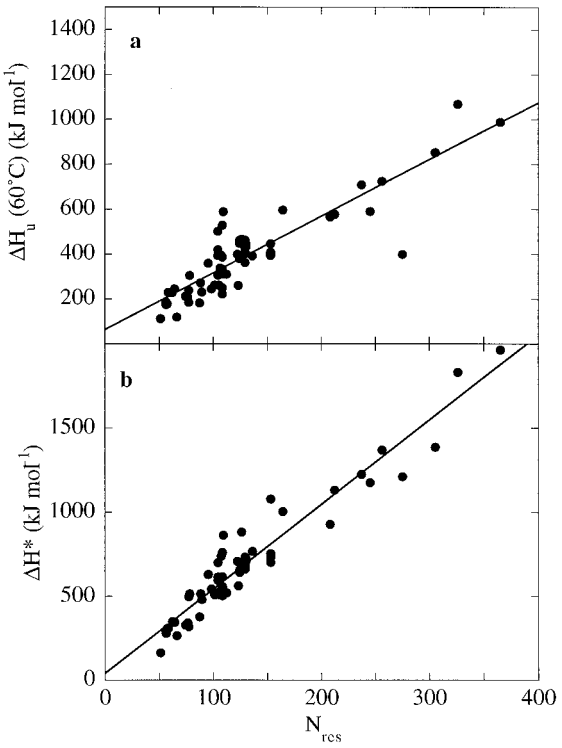

Figure 5. Correlation of ΔH_u with the number of residues at 60 °C (a) and 100.5 °C (b). The lines are the linear regressions. The slope, intercept, and R^2 values are 2.53, 63.2, and 0.789 at 60 °C, and 5.03, 41.6, and 0.922 at 100.5 °C.

al.[27] gives $T_H^* = 84$ °C and $\Delta H^* = 4.62$ kJ (mol res)$^{-1}$.

The linear regression of the ΔS_u data gives a slope of −0.126 and an intercept of 10.1 with $R^2 = 0.33$. This corresponds to $T_S^* = 64.9$ °C and $\Delta S^* = 10.1$ J K^{-1} (mol res)$^{-1}$. The dotted line again corresponds to the analysis of Murphy and Gill[34] with $T_S^* = 112$ °C and $\Delta S^* = 18.1$ J K^{-1} (mol res)$^{-1}$.

Overall, the convergence behavior for this larger protein data is not very compelling. This is not surprising given the above discussion. The correlation coefficients suggest that only about 30% of the variation in ΔH_u or ΔS_u of unfolding at 25 °C can be accounted for by variation in ΔC_p. In contrast, regression coefficients from the data set of 12 proteins originally analyzed by Murphy and Gill suggest that the variation in ΔC_p accounts for >90% of the variation in ΔH_u and ΔS_u of unfolding at 25 °C.

4. Enthalpy of Unfolding

The ΔH_u at 60 °C can also be treated as a function of N_{res}, ΔA_{tot}, or ΔA_{ap} and ΔA_{pol}. Regression on N_{res} (Figure 5a) yields 2.92 ± 0.08 kJ (mol res)$^{-1}$ with $R^2 = 0.766$, while regression on ΔA_{tot} yields 30.2 ± 0.9 J (mol Å^2)$^{-1}$ with $R^2 = 0.735$ (Table 5).

As with ΔC_p, results of model compound studies show that apolar and polar ASA make different contributions to ΔH_u.[27,30,35,79,86,87] Regression of the ΔH_u of unfolding at 60 °C on ΔA_{ap} and ΔA_{pol} yields values of −8 ± 11 and 86 ± 17 J (mol Å^2)$^{-1}$ respectively with $R^2 = 0.775$. While the regression in terms of both ΔA_{ap} and ΔA_{pol} is statistically better than for ΔA_{tot}, the confidence in the regressed parameters is much less.

Analysis of a smaller protein data set by Xie and Freire[88] gave values of −35.3 J (mol Å^2)$^{-1}$ for apolar surface and 131 J (mol Å^2)$^{-1}$ for polar surface, in reasonable agreement with the results of our larger data set. Values of −21.5 J (mol Å^2)$^{-1}$ for apolar surface and 205 J (mol Å^2)$^{-1}$ are calculated from data on the dissolution of cyclic dipeptides.[12,79]

As discussed above, the apolar and polar contributions to ΔH_u have also been estimated from the convergence temperatures.[33,80,85] In this analysis, it is assumed that the apolar contribution to ΔH_u is zero at T_H^*. The ΔH_u observed at that temperature, ΔH^*, can then be normalized to the change in polar surface area to give the polar contribution to ΔH_u.

From the data set compiled by Privalov and Gill,[38] Murphy and Gill determined a T_H^* of 100.5 °C at which ΔH^* equals 5.64 kJ (mol res)$^{-1}$.[34] This corresponds to 146 J (mol Å^2)$^{-1}$ when normalized to surface area.[35] From this and the ΔC_p values used in the convergence model[35] the calculated contributions at 60 °C are −76 J (mol Å^2)$^{-1}$ for apolar surface and 190 J (mol Å^2)$^{-1}$ for polar surface.

The current data set yields a T_H^* of 65.9 °C. However, if the ΔH_u values are extrapolated to the original T_H^* of 100.5 °C, the correlation between ΔH_u and N_{res} is much improved (Figure 5b), with a value of 5.28 ± 0.09 kJ (mol res)$^{-1}$ and $R^2 = 0.919$. Thus, even though convergence behavior is not very evident in the protein data set, the value of ΔH_u is well predicted at the original T_H^* of 100.5 °C.

As with ΔC_p we find that all of the analyses give values for the average ΔH_u per unit surface area at 60 °C, when weighted by the percentage of apolar and polar surface, that are very similar. The values from the "convergence temperature" analysis give an average ΔH_u of 34.9 J (mol Å^2)$^{-1}$; the values of Xie and Freire give 34.0 J (mol Å^2)$^{-1}$; and the values from the current analysis give 31.2 J (mol Å^2)$^{-1}$. These compare well with the value of 30.2 J (mol Å^2)$^{-1}$ obtained when the current data set is regressed against ΔA_{tot}. Overall, this comparison again illustrates the difficulty of obtaining precise coefficients from the protein data alone.

5. Entropy of Unfolding

Both ΔC_p and ΔH_u are expected to scale with changes in accessible surface area because these quantities result primarily from changes in solvation and changes in noncovalent interactions. The entropy change, on the other hand, includes additional contributions from changes in the configurational entropy of side chains and backbone upon unfolding. Regression of the ΔS_u at 60 °C on the number of residues (Figure 6a) gives a value of 8.8 ± 3 J K^{-1} (mol res)$^{-1}$ with $R^2 = 0.744$. The correlation of ΔS_u at 60 °C with the total buried surface area is somewhat poorer with a value of 0.091 ± 0.003 J K^{-1} (mol Å^2)$^{-1}$ and $R^2 = 0.716$. The best correlation of ΔS_u at 60 °C is with both ΔA_{ap} and ΔA_{pol}, giving values of −0.03 ± 0.04 J K^{-1} (mol Å^2)$^{-1}$ and 0.27 ± 0.06 J K^{-1} (mol Å^2)$^{-1}$ with $R^2 = 0.757$ (Table 5).

Upon the basis of model compound data, one would expect the apolar contribution to ΔS_u at 60 °C to be about −0.2 to −0.3 J K^{-1} (mol Å^2)$^{-1}$. This value is based on the typical ΔC_p and $T_S^* = 112$ °C. The magnitude of regressed value is significantly less than this. The estimated polar contribution to ΔS_u is small[85,89] or negative,[6] but the regressed value is

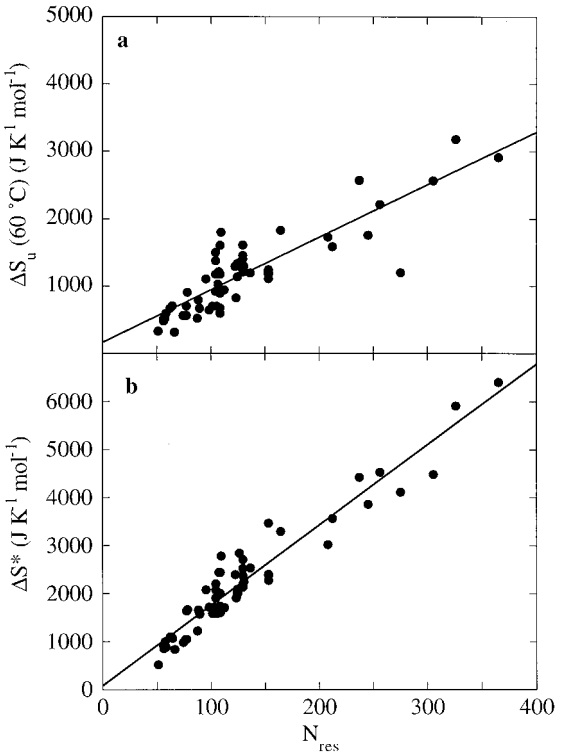

Figure 6. Correlation of ΔS_u with the number of residues at 60 °C (a) and 112 °C (b). The lines are the linear regressions. The slope, intercept, and R^2 values are 7.8, 162, and 0.759 at 60 °C, and 16.8, 85, and 0.920 at 112 °C.

large and positive. The discrepancies between the regressed values for the ΔS_u contributions of apolar and polar surface and the expectation from model compound and theoretical studies is due to the neglect of the configurational entropy in the regression analysis. If a regression is performed against the number of residues and the changes in apolar and polar surface areas, the resulting coefficients are 9.2 $\pm$ 4.6 J K^{-1} (mol res)$^{-1}$, -0.11 ± 0.5 J K^{-1} (mol Å^2)$^{-1}$, and 0.15 ± 0.08 J K^{-1} (mol Å^2)$^{-1}$ respectively with $R^2 = 0.771$. These values are in much better agreement with expectation.

As with ΔH_u, the convergence behavior of ΔS_u for this set of proteins is quite weak, but gives a convergence temperature of 64.9 °C. However, if the values of ΔS_u are extrapolated to the more commonly observed value of $T_S^* = 112$ °C, a much improved correlation with the number of residues is again observed (Figure 6b), giving 17.3 ± 0.3 J K^{-1} (mol res)$^{-1}$ with $R^2 = 0.919$. This is very similar the value of 18.1 J K^{-1} (mol res)$^{-1}$ observed in the smaller data set.[33]

6. Comparison of Regressed and Experimental Values

Coefficients obtained from the regressions described above (Table 5) can be used to calculate the thermodynamics of unfolding, which can then be compared to the experimental values. The regression parameters describe the average thermodynamics of these proteins, so comparisons of the calculated and experimental values provide some information on how much a protein deviates from the average behavior of globular proteins.

The ΔC_p values are calculated using the parameter for ΔA_{tot} given in Table 5. The calculated values and the percentage error are given in Table 6. In this

Table 6. Comparison of Calculated and Experimental Values of ΔC_p[a]

protein	calculated ΔC_p	% error
α-chymotrypsin	13.6	6.6
α-chymotrypsinogen	14.0	−3.4
α-lactalbumin	7.1	−5.2
α-lactalbumin	7.0	−7.9
acyl carrier protein (apo)	3.7	10.6
acyl carrier protein (holo)	3.7	−42.6
arabinose binding protein	19.1	44.6
arc repressor	6.1	−8.3
B1 of protein G	2.8	8.4
B2 of protein G	3.1	6.4
barnase	6.4	10.3
barnase	6.3	−7.2
barstar	5.1	−18.9
BPTI	2.8	41.4
carbonic anhydrase B	16.1	0.7
CI2	3.4	34.8
cyt *b*5 (tryp frag)	4.5	−24.8
cytochrome *c* (horse)	5.8	15.1
cytochrome *c* (horse)	5.8	7.8
cytochrome *c* (yeast iso 1)	5.9	3.5
cytochrome *c* (yeast iso 1)	5.9	13.4
cytochrome *c* (yeast iso 2)	6.0	15.8
GCN4	3.2	8.5
HPr	4.6	−5.3
IL-1β	8.5	5.8
lac repressor headpiece	2.4	82.2
lysozyme (human)	7.8	9.0
lysozyme (human)	7.8	18.1
lysozyme (apo equine)	7.5	18.6
lysozyme (holo equine)	7.7	1.3
lysozyme (hen)	7.7	4.0
lysozyme (hen)	7.5	16.9
lysozyme (hen)	7.5	11.7
lysozyme T4	9.9	−1.6
met repressor	12.4	39.1
myoglobin (horse)	8.7	14.3
myoglobin (whale)	9.0	−42.6
myoglobin (whale)	9.0	2.4
OMTKY3	2.4	−5.6
papain	13.2	−3.8
parvalbumin	5.9	5.5
pepsin	19.0	0.6
pepsinogen	22.5	−6.8
plasminogen K4 domain	4.4	−16.4
RNase T1	5.4	10.8
RNase T1	5.4	11.1
RNaseA	6.8	4.0
RNaseA	6.8	42.2
RNaseA	6.8	41.9
ROP	7.8	−24.0
Sac7d	3.6	−1.1
SH3 spectrin	3.2	−1.7
Staphylococcus nuclease	8.0	−13.5
stefin A	5.3	−28.4
stefin B	5.3	−21.2
subtilisin inhibitor	5.2	−38.8
subtilisin BPN′	15.7	−21.7
tendamistat	3.7	28.3
thioredoxin	5.9	−14.8
thioredoxin	5.9	−19.5
trp repressor[b]	6.2	1.9
ubiquitin	4.1	22.1

[a] ΔC_p (kJ K^{-1} mol^{-1}) values were calculated as a function of N_{res} using the regression coefficients listed in Table 5. Errors are calculated in comparison to the experimental values in Table 2 as 100 × (calculated − experimental)/experimental. [b] Using the 2WRP structure.

and subsequent tables, the structural data are taken from Table 3, and the calculations are compared to the experimental values in Table 2. In cases where there are multiple thermodynamic entries in Table 2, but a single structural entry in Table 3, the

calculated values are compared to each of the experimental entries. In those cases where multiple structure and thermodynamic entries are given, the comparison is made between structures and thermodynamics in the same order in which they are given in Tables 2 and 3. The percentage error is calculated as 100 × (calculated − experimental)/experimental. The average error in calculating ΔC_p is 4 ± 22%. The average is expected to be small as overpredictions and underpredictions cancel each other, but the standard deviation indicates that the error in the prediction is larger than the estimated experimental error.

The ΔH_u values at 60 °C are calculated using the parameters for both ΔA_{ap} and ΔA_{pol} given in Table 5 and are summarized in Table 7. The average error is again small, −2.8 ± 22%, but the standard deviation is large. Table 7 also lists the error in calculating ΔH^* (at $T_H^* = 100$ °C) which has an average error of 2 ± 16%. Thus, as evident in the regression coefficients, ΔH^* is better predicted than ΔH_u at 60 °C.

Finally, the ΔS_u values at 60 °C are calculated using the parameters for N_{res}, ΔA_{ap}, and ΔA_{pol} given in Table 5 and are summarized in Table 8. The average error is 5 ± 26%. The calculated values of ΔS^* (at $T_S^* = 112$ °C) are also given in the table and have an average error of 2 ± 17%. Again, the values of ΔS^* are better predicted than the values of ΔS_u at 60 °C.

One possible explanation for error in the predictions is deviations from the mean structural characteristics of the proteins. For example, greater numbers of disulfide bonds are expected to lead to decreases in ΔS_u, so that one might expect ΔS_u to be overpredicted for proteins with a greater than average number of disulfides. In fact, no such correlation is seen between the number of disulfides and either ΔS_u at 60 °C or ΔS^* (Figure 7). The correlation coefficients, R^2, are less than 0.05 for both cases. In fact, no correlation of the error in either ΔS_u at 60 °C or ΔS^* is observed with any of the structural features considered here, including the fraction of the buried surface area which is polar or apolar, the percentage of the residues in any secondary structure type (i.e., α-helix, β-sheet, β-turn, or the sum of all three), and the number of residues. There is also no correlation with the experimental parameters such as pH or T_m.

The same lack of correlation of the error in prediction with any structural or experimental features is observed for ΔH_u at 60°C, ΔH^*, and ΔC_p. It is somewhat surprising that no correlation of error in predicting ΔH_u is found with the percentage of residues in any secondary structural type as such a correlation has previously been noted for a smaller data set.[90] In fact, the only significant correlation we have observed is between the error in ΔH_u and the error in ΔS_u. This is illustrated in Figure 8a. The line is the linear least-squares fit which has a slope of 1 and an intercept of 7 with $R^2 = 0.756$. This correlation is even more evident between the error in ΔH^* and the error in ΔS^*, as seen in Figure 8b in which the slope is 1, the intercept is 0.4 and $R^2 = 0.926$.

Table 7. Comparison of Calculated and Experimental Values of ΔH_u[a]

name of protein	ΔH_u (60 °C)	error, %	ΔH^*	error, %
α-chymotrypsin	640	−9.7	1252	2.2
α-chymotrypsinogen	680	15.3	1294	9.9
α-lactalbumin	353	35.9	650	15.3
α-lactalbumin	363	−9.1	645	−9.0
acyl carrier protein (apo)	212	14.9	407	27.2
acyl carrier protein (holo)	212	−10.9	407	−18.5
arabinose binding protein	901	5.6	1611	16.1
arc repressor	358	6.1	560	−7.9
B1 of protein G	147	−21.2	296	1.4
B2 of protein G	160	−12.1	296	−1.1
barnase	326	−38.3	571	−25.1
barnase	322	−45.3	576	−33.4
barstar	202	−12.1	470	−2.6
BPTI	148	−35.4	306	−1.2
carbonic anhydrase B	792	9.2	1352	−1.4
CI2	164	−33.3	338	−2.6
cyt *b*5 (tryp frag)	235	−13.6	465	−9.7
cytochrome *c* (horse)	283	−28.0	549	−7.8
cytochrome *c* (horse)	283	−7.8	549	5.0
cytochrome *c* (yeast iso 1)	308	−20.3	571	−7.6
cytochrome *c* (yeast iso 1)	308	−1.3	571	9.1
cytochrome *c* (yeast iso 2)	330	6.1	592	13.5
GCN4	181	−21.0	328	−6.3
HPr	227	24.2	460	21.2
IL-1β	378	−7.1	808	10.6
lac repressor headpiece	122	9.8	269	64.1
lysozyme (human)	423	−2.6	687	−5.1
lysozyme (human)	423	−4.9	687	−3.5
lysozyme (apo equine)	425	5.9	682	−3.9
lysozyme (holo equine)	425	17.7	682	3.1
lysozyme (hen)	405	−5.1	682	0.0
lysozyme (hen)	405	−1.0	682	2.1
lysozyme (hen)	405	−12.5	682	−7.1
lysozyme T4	505	−15.3	866	−13.7
met repressor	642	13.4	1099	18.4
myoglobin (horse)	408	3.7	808	15.0
myoglobin (whale)	443	−0.7	808	−25.1
myoglobin (whale)	420	5.3	808	7.2
OMTKY3	145	−17.0	296	5.7
papain	650	12.5	1120	−1.1
parvalbumin	302	−9.2	571	2.1
pepsin	861	−19.5	1722	−6.0
pepsinogen	1059	7.0	1928	−1.9
plasminogen K4 domain	265	−13.0	412	−20.1
RNase T1	292	−30.4	549	−10.8
RNase T1	290	−42.3	549	−21.4
RNaseA	427	12.8	655	1.6
RNaseA	427	−7.5	655	−0.2
RNaseA	427	−4.6	655	1.9
ROP	534	14.4	666	−24.7
Sac7d	191	58.9	349	31.3
SH3 spectrin	147	−17.2	301	−2.6
Staphylococcus nuclease	385	−1.9	718	−6.3
stefin A	274	12.0	518	−5.0
stefin B	263	−26.7	502	−20.3
subtilisin inhibitor	270	−31.8	565	−23.4
subtilisin BPN′	769	92.5	1453	19.7
tendamistat	215	1.4	391	18.9
thioredoxin	250	12.8	571	13.3
thioredoxin	250	0.4	571	4.2
trp repressor	309	17.4	555	8.8
ubiquitin	193	−7.1	402	17.1

[a] ΔH_u (kJ mol^{-1}) at 60 °C was calculated as a function of ΔA_{ap} and ΔA_{pol}, and ΔH^* (i.e., ΔH at 100 °C) was calculated as a function of N_{res}, using the regression coefficients in Table 5. Errors are calculated in comparison to the experimental values in Table 2.

IV. Summary

What conclusions can be drawn from the regression analyses? The first conclusion is that, from a purely empirical standpoint, the primary determinant of

Table 8. Comparison of Calculated and Experimental ΔS_u[a]

name of protein	ΔS_u (60 °C)	error, %	ΔS^*	error, %
α-chymotrypsin	1985	−22.9	4095	−7.4
α-chymotrypsinogen	2108	19.8	4233	9.6
α-lactalbumin	1079	31.0	2125	11.2
α-lactalbumin	1111	−14.0	2108	−12.1
acyl carrier protein (apo)	759	34.1	1330	26.7
acyl carrier protein (holo)	759	7.6	1330	−18.8
arabinose binding protein	2537	−1.2	5270	17.6
arc repressor	1074	4.3	1831	−8.4
B1 of protein G	513	0.8	968	9.2
B2 of protein G	510	−0.2	968	3.9
barnase	973	−39.5	1866	−23.7
barnase	981	−45.5	1883	−32.4
barstar	650	−2.8	1538	−2.2
BPTI	533	−10.1	1002	13.6
carbonic anhydrase B	2220	0.1	4423	−2.4
CI2	554	−21.6	1106	3.4
cyt *b*5 (tryp frag)	806	2.0	1520	−8.4
cytochrome *c* (horse)	991	−15.7	1797	−5.7
cytochrome *c* (horse)	907	−23.2	1797	5.9
cytochrome *c* (yeast iso 1)	907	−1.6	1866	−6.8
cytochrome *c* (yeast iso 1)	991	4.6	1866	9.6
cytochrome *c* (yeast iso 2)	1068	12.8	1935	13.7
GCN4	606	−9.3	1071	−2.4
HPr	763	45.7	1503	22.5
IL-1β	1230	−1.3	2644	9.9
lac repressor headpiece	816	−10.2	881	70.1
lysozyme (human)	464	40.8	2246	−0.3
lysozyme (human)	1235	1.6	2246	−0.4
lysozyme (apo equine)	1235	−4.8	2229	−17.8
lysozyme (holo equine)	1223	−4.5	2229	−11.7
lysozyme (hen)	1247	−22.5	2229	1.6
lysozyme (hen)	1247	−14.1	2229	4.2
lysozyme (hen)	1223	0.9	2229	−6.3
lysozyme T4	1223	−13.1	2834	−14.1
met repressor	1465	−20.0	3594	18.7
myoglobin (horse)	1888	9.0	2644	15.9
myoglobin (whale)	1276	20.8	2644	−23.8
myoglobin (whale)	1337	10.7	2644	10.8
OMTKY3	1271	14.1	968	12.9
papain	561	16.6	3663	2.5
parvalbumin	1841	16.0	1866	9.4
pepsin	972	8.7	5633	−4.7
pepsinogen	2641	−16.9	6306	−1.6
plasminogen K4 domain	3038	4.3	1348	−19.1
RNase T1	1331	16.7	1797	−13.6
RNase T1	984	−28.5	1797	−18.6
RNaseA	973	−35.2	2142	2.4
RNaseA	1331	2.2	2142	7.2
RNaseA	1331	−0.5	2142	5.3
ROP	1496	10.8	2177	−23.5
Sac7d	620	96.1	1140	36.3
SH3 spectrin	469	−10.4	985	−0.9
Staphylococcus nuclease	1157	−3.4	2350	−7.5
stefin A	893	38.3	1693	−1.5
stefin B	836	−24.4	1641	−21.0
subtilisin inhibitor	2381	97.6	1849	−24.4
subtilisin BPN′	981	−19.5	4751	15.3
tendamistat	736	30.2	1279	29.8
thioredoxin	831	39.4	1866	16.3
thioredoxin	831	23.5	1866	7.2
trp repressor	920	31.2	1814	14.4
ubiquitin	645	15.0	1313	25.8

[a] ΔS_u (J K^{-1} mol^{-1}) at 60 °C was calculated as a function of ΔA_{ap} and ΔA_{pol}, and ΔS^* (i.e., ΔS_u at 112 °C) was calculated as a function of N_{res}, using the regression coefficients in Table 5. Errors are calculated in comparison to the experimental values in Table 2.

protein unfolding thermodynamics is the size of the protein, although the number of residues by itself does not give the best regression. The regression analysis indicates also that at least 75% of the variation in unfolding energetics can be accounted for by the variation in simple structural features such as buried surface area. Thus, this simple approach

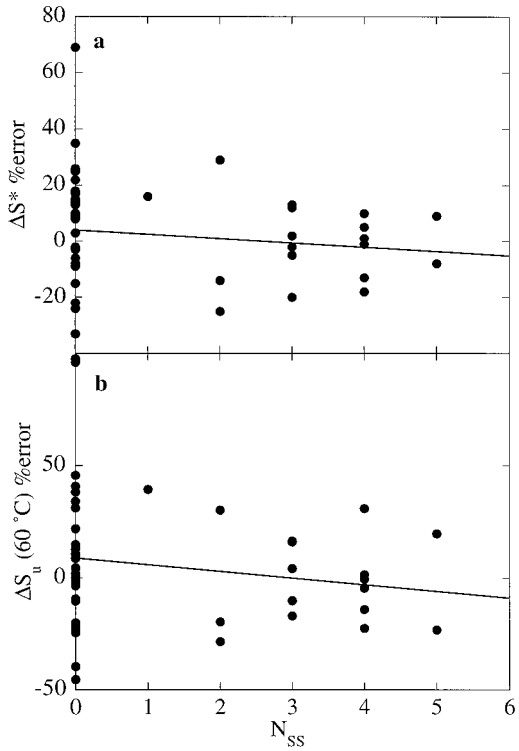

Figure 7. Correlation of the percentage error in calculating ΔS_u at 112 °C (a) and 60 °C (b) with the number of disulfide bonds in the protein. The lines are the linear regressions. The slope, intercept, and R^2 are −1.5, 4.0, and 0.021 at 112 °C, and −3.0, 8.8, and 0.032 at 60 °C. There is no improvement in the correlation if the proteins with zero disulfides are excluded from analysis.

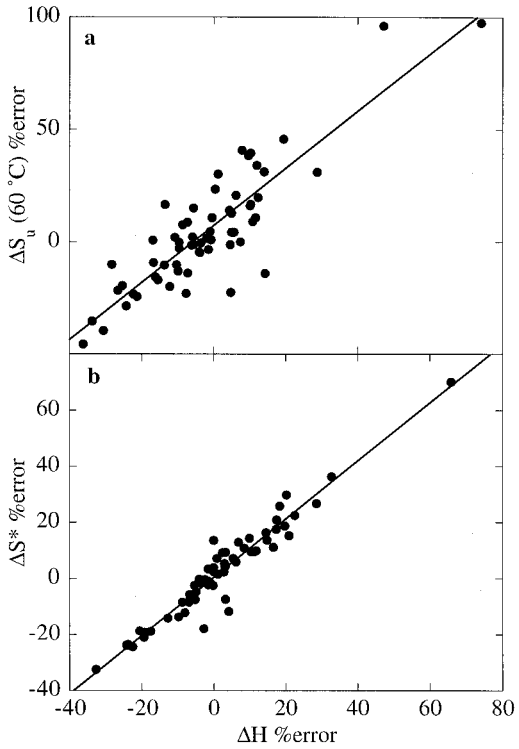

Figure 8. Correlation of the percentage error in calculating ΔS_u and ΔH_u. (a) At 60 °C, the line is the linear regression with slope, intercept, and R^2 of 1.0, 7.3, and 0.756. (b) ΔS_u at 112 °C and ΔH_u at 100.5 °C, the line is the linear regression. The slope, intercept, and R^2 are 1.0, 0.4, and 0.927.

captures most of the important features which determine protein energetics. Further evidence in favor of analyses based on surface areas is the recent work of Hilser and Freire:[91] calculations of protein energetics based on surface areas were successfully used to predict amide hydrogen exchange behavior in a number of proteins.

It is also interesting to note that the thermodynamics at the convergence temperatures observed previously in a much smaller set of proteins are better predicted than the thermodynamics at 60 °C, even though the current set of proteins do not convincingly show convergence behavior. This is perhaps not surprising for the entropy, as the value of T_S* seems to be fairly universal.[33] It is not clear why it should also occur for T_H*.

Calculated values of ΔC_p range from 57% to 182% of the experimental values (Table 6). The range for ΔH_u at 60 °C is 55% to 159% of the experimental values while that for ΔH* is 67% to 164% (Table 7). Similar distributions are observed in the differences between calculated and experimental values of ΔS_m (Table 8). The extent to which calculated values of ΔH_m and ΔS_m are under- or overestimated relative to experimental values is highly correlated, which probably reflects the fact that experimental ΔH_u values are used to calculate ΔS_u values: experimental errors in ΔH_u are thus manifested in the relative errors in ΔS_u. Interestingly, distributions of differences between calculated and experimental values are broader, $\pm$16−27% at one standard deviation, than might be anticipated on the basis of experimental error alone, which is about 10% on average (7% for ΔC_p, 12% for $\Delta H(60)$, and 15% for $\Delta S(60)$ as noted above in section II.C).

A likely explanation for the broad distribution in the differences between the calculated and observed parameters is inaccuracies in the model used in the regression analysis, which is based primarily on surface area differences. Moreover, the calculations rely on convergence temperatures and the protein data show considerable scatter in this regard (Figure 6). Overall, empirical correlations of energetics with "regular" features of protein structure give rise to errors that appear to exceed experimental error. The model is thus either inappropriate or incomplete. Because the model based on surface areas appears to capture much, but not all, of the relationship between protein structure and the energetics of protein stability, the simplest explanation is that the model is incomplete. Inclusion of information about secondary structure and disulfide bonds provides no insight into the origin of discrepancies between calculated and observed energetics.

What is missed when the energetics of protein stability are decomposed in terms of changes in solvent-exposed surface areas? Some possible answers to this question are (1) nonadditivity of energetic contributions from the various groups that make up polar and nonpolar surfaces, (2) long-range interactions in proteins, and (3) heterogeneity in the extent to which the denatured states for different proteins are exposed to solvent. The possibility of nonadditivity in protein energetics is the subject of considerable discussion.[7,92,93] The principle of additivity is that the observed thermodynamics of protein stability result from a simple sum of independent contributions from individual interactions. Deconvolution of protein stability in terms of polar and nonpolar surface areas is predicated on the assumption that the contributions from such surfaces are linear functions of surface area. Nonadditivity may well contribute to the scatter in the calculated vs observed energetics, but no straightforward approach is available yet for evaluating its role.

Electrostatic interactions are the principal long-range interactions in proteins. With a database consisting of many different proteins, differences in the extent to which electrostatic interactions contribute to stability in different proteins are going to contribute to the error in parameters derived from regression analysis of the database.

No direct experimental data are available for assessing the amount of new surface area that is exposed when a protein unfolds. Analyses of protein stability with respect to solvent-exposed surface areas typically rely on the assumption that solvent exposure in the denatured state is modeled accurately by an extended polypeptide chain or by summing calculated surface areas for tripeptides.[26,94] The use of different algorithms for these calculations leads to significant differences in surface areas, but use of a single algorithm in deconvoluting energetics in terms of structure is only expected to lead to *systematic* deviations with respect to results obtained using other algorithms.[69] If, however, proteins differ in the extent to which their denatured states are exposed to solvent, then considerable error will be introduced into the analysis regardless of the algorithm used to calculate surface area.

A number of investigators have argued that the denatured state is not accurately modeled by an extended or random-coil polypeptide chain.[41,42,95−99] Moreover, the extent of solvent exposure is proposed to be sensitive to solution conditions, so no one value for solvent-exposed surface area in the denatured state is applicable to any protein. Is the proposed heterogeneity in the extent of unfolding a reasonable explanation for the disagreement between calculated and experimental values for the energetics of protein stability?

One intriguing observation in this regard is the underestimated ΔH_u and ΔS_u values for barnase and RNase T1 (Table 7); these two proteins fall at the extreme low end for both parameters. Confidence in the experimental determinations is high because at least two independent determinations have been made for each protein. Interestingly, barnase and RNase T1 have very similar three-dimensional structures in spite of the fact that their amino acid sequences are only 14% identical. Finally, the extent of unfolding in the denatured state of barnase appears to be high relative to other proteins.[100]

If the extent of solvent exposure for the denatured states of barnase and RNase T1 is indeed greater than the average for all proteins in the database then one would expect that, for barnase and RNase T1, the thermodynamic parameters calculated from the mean behavior for all proteins would be lower than the true values, as is observed (Table 7). However, this behavior is not observed for ΔC_p (Table 6). In addition, the extent of unfolding for RNase T1 ap-

pears to be close to that for other proteins.[95,100] Nevertheless, at least some of the experimental data tabulated here and presented elsewhere are consistent with variability in the extent of unfolding for different proteins.

The possibility that the denatured state of barnase is more unfolded than the average protein suggests that the average extent of unfolding for all proteins is overestimated with the current algorithms. A low estimate for the average extent of unfolding can be obtained by using barnase as a reference for denatured protein that is completely exposed to solvent. In conjunction with the observation that calculated ΔH_u and ΔS_u values are $\leq$75% of the predicted values (Tables 7 and 8), this suggests that the average extent of unfolding is $\leq$75% of the values calculated with the model-based algorithms. This value is similar to those suggested by Lee[82] and Brandts[101] and is consistent with the conclusions of a recent computational study,[69] where alternative models for the denatured state yielded surface areas that averaged about 80% of the values obtained with tripeptides.

An important conclusion from this analysis is that additional refinement of the calculations and a molecular interpretation of the regression coefficients are unlikely to come from the protein data themselves. The inability to obtain unique coefficients which relate structural features to unfolding energetics may reflect variability in the quality of the data or variability in the validity of the assumptions across the data set; the latter appears to be likely. Rather than simply compile additional protein unfolding thermodynamics for a wide variety of proteins, it may be more promising to pursue systematic structural and calorimetric studies of single-site mutations or structurally homologous proteins. The idea here is that the differences between the proteins in such studies would more closely conform to those of a homologous series.

More data concerning the denatured state are essential for progress in understanding the energetics of protein stability. In this regard, calorimetric experiments appear to offer some promise.[39,69] Additionally, data on protein−protein interactions,[102] in which the structures of both the initial and final states are well determined, will probably provide less ambiguous regression values. Finally, model compound studies will continue to be the principal means by which precise thermodynamic values for specific interactions can be determined. These studies provide a rich framework to guide design and interpretation of the protein studies.

V. Acknowledgments

The authors thank the reviewers and Professor Ken A. Dill for critical reading and helpful comments. We also thank Dr. Wesley Stites for providing a copy of his contribution to this volume prior to publication. The authors are grateful to the National Institutes of Health, National Science Foundation, American Chemical Society−Petroleum Research Fund, and the University of Iowa for support of this work.

VI. References

(1) Anfinsen, C. B. *Science* **1973**, *181*, 223.
(2) *The Biology of Nonspecific DNA-Protein Interactions*; Revzin, A., Ed.; CRC Press: Boca Raton, 1990.
(3) Lattman, E. E.; Rose, G. D. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 439.
(4) Yue, K.; Dill, K. A. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 146.
(5) Kauzmann, W. *Adv. Protein Chem.* **1959**, *14*, 1.
(6) Makhatadze, G. I.; Privalov, P. L. *Adv. Protein Chem.* **1995**, *47*, 307.
(7) Lazaridis, T.; Archontis, G.; Karplus, M. *Adv. Protein Chem.* **1995**, *47*, 231.
(8) Honig, B.; Yang, A.-S. *Adv. Protein Chem.* **1995**, *46*, 27.
(9) Rose, G. D.; Wolfenden, R. *Annu. Rev. Biophys. Biomol. Struct.* **1993**, *22*, 381.
(10) Creighton, T. E. *Curr. Opin. Struct. Biol.* **1991**, *1*, 5.
(11) Dill, K. A. *Biochemistry* **1990**, *29*, 7133.
(12) Habermann, S. M.; Murphy, K. P. *Protein Sci.* **1996**, *5*, 1229.
(13) Baldwin, R. L. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 8069.
(14) Dill, K. A. *Science* **1990**, *250*, 297.
(15) Herzfeld, J. *Science* **1991**, *253*, 88.
(16) Privalov, P. L.; Gill, S. J.; Murphy, K. P. *Science* **1990**, *250*, 297.
(17) Chothia, C. *Nature* **1975**, *254*, 304.
(18) Myers, J. K.; Pace, C. N. *Biophys. J.* **1996**, *71*, 2033.
(19) Levitt, M.; Chothia, C. *Nature* **1976**, *261*, 552.
(20) Chothia, C. *J. Mol. Biol.* **1976**, *105*, 1.
(21) Richardson, J. *Adv. Protein Chem.* **1981**, *34*, 167.
(22) Chothia, C. *Annu. Rev. Biochem.* **1984**, *53*, 537.
(23) Thornton, J. M. In *Protein Folding*; Creighton, T. E., Ed.; W. H. Freeman and Co.: New York, 1992; pp 59.
(24) Privalov, P. L. *Adv. Protein Chem.* **1979**, *33*, 167.
(25) Spolar, R. S.; Ha, J.-H.; Record, M. T., Jr. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 8382.
(26) Livingstone, J. R.; Spolar, R. S.; Record, M. T., Jr. *Biochemistry* **1991**, *30*, 4237.
(27) Spolar, R. S.; Livingstone, J. R.; Record, M. T., Jr. *Biochemistry* **1992**, *31*, 3947.
(28) Nozaki, Y.; Tanford, C. *J. Biol. Chem.* **1971**, *246*, 2211.
(29) Privalov, P. L.; Makhatadze, G. I. *J. Mol. Biol.* **1992**, *224*, 715.
(30) Makhatadze, G. I.; Privalov, P. L. *J. Mol. Biol.* **1993**, *232*, 639.
(31) Privalov, P. L.; Makhatadze, G. I. *J. Mol. Biol.* **1993**, *232*, 660.
(32) Murphy, K. P.; Gill, S. J. *J. Chem. Thermodyn.* **1989**, *21*, 903.
(33) Murphy, K. P.; Privalov, P. L.; Gill, S. J. *Science* **1990**, *247*, 559.
(34) Murphy, K. P.; Gill, S. J. *J. Mol. Biol.* **1991**, *222*, 699.
(35) Murphy, K. P.; Freire, E. *Adv. Protein Chem.* **1992**, *43*, 313.
(36) Ooi, T.; Oobatake, M.; Némethy, G.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 3086.
(37) Eisenberg, D.; McLachlan, A. D. *Nature* **1986**, *319*, 199.
(38) Privalov, P. L.; Gill, S. J. *Adv. Protein Chem.* **1988**, *39*, 191.
(39) Privalov, P. L.; Tiktopulo, E. I.; Yenyaminov, S. Y.; Griko, Y. V.; Makhatadze, G. I.; Khechinashvili, N. N. *J. Mol. Biol.* **1989**, *205*, 727.
(40) Robertson, A. D.; Baldwin, R. L. *Biochemistry* **1991**, *30*, 9907.
(41) Dill, K. A.; Shortle, D. *Annu. Rev. Biochem.* **1991**, *60*, 795.
(42) Shortle, D. *FASEB J.* **1996**, *10*, 27.
(43) Edsall, J. T. *J. Am. Chem. Soc.* **1935**, *57*, 1506.
(44) Madan, B.; Sharp, K. *J. Phys. Chem.* **1996**, *100*, 7713.
(45) Gill, S. J.; Dec, S. F.; Olofsson, G.; Wadsö, I. *J. Phys. Chem* **1985**, *89*, 3758.
(46) Privalov, P. L.; Potekhin, S. A. *Methods Enzymol.* **1986**, *131*, 4.
(47) Freire, E. In *Protein Stability and Folding*; Shirley, B. A., Ed.; Humana Press: Totowa, NJ, 1995; Vol. 40, pp 191.
(48) Christensen, J. J.; Hansen, L. D.; Izatt, R. M. *Handbook of Proton Ionization Heats and Related Thermodynamic Quantities*; John Wiley and Sons: New York, 1976.
(49) Freire, E.; Biltonen, R. L. *Biopolymers* **1978**, *17*, 463.
(50) Bowie, J. U.; Sauer, R. T. *Biochemistry* **1989**, *28*, 7139.
(51) Swint, L.; Robertson, A. D. *Protein Sci.* **1993**, *2*, 2037.
(52) Cohen, D. S.; Pielak, G. J. *Protein Sci.* **1994**, *3*, 1253.
(53) Scholtz, J. M. *Protein Sci.* **1995**, *4*, 35.
(54) Pace, C. N.; Laurents, D. V. *Biochemistry* **1989**, *28*, 2520.
(55) Chen, B.-l.; Schellman, J. A. *Biochemistry* **1989**, *28*, 685.
(56) Santoro, M. M.; Bolen, D. W. *Biochemistry* **1988**, *27*, 8063.
(57) Kim, P. S.; Baldwin, R. L. *Annu. Rev. Biochem.* **1982**, *51*, 459.
(58) Pace, C. N.; Vajdos, F.; Fee, L.; Grimsley, G.; Gray, T. *Protein Sci.* **1995**, *4*, 2411.
(59) Becktel, W. J.; Schellman, J. A. *Biopolymers* **1987**, *26*, 1859.
(60) Carra, J. H.; Anderson, E. A.; Privalov, P. L. *Protein Sci.* **1994**, *3*, 944.
(61) DeKoster, G. T.; Robertson, A. D. *Biochemistry* **1997**, *36*, 2323.
(62) Bevington, P. R.; Robinson, D. K. *Data Reduction and Error Analysis for the Physical Sciences*, 2nd ed.; McGraw-Hill: New York, 1992; pp 328.
(63) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E. F., Jr.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. *J. Mol. Biol.* **1977**, *112*, 535.
(64) Abola, E. E.; Bernstein, F. C.; Bryant, S. H.; Koetzle, T. F.; Weng, J. In *Crystallographic Databases - Information Content, Software Systems, Scientific Applications*; Allen, F. H., Bergerhoff, G., Sievers, R., Eds.; Data Commission of the International Union of Crystallography: Bonn/Cambridge/Chester, 1987; p 107.
(65) Lee, B.; Richards, F. M. *J. Mol. Biol.* **1971**, *55*, 379.
(66) Russell, R. B.; Barton, G. J. *J. Mol. Biol.* **1994**, *244*, 332.

(67) Flores, T. P.; Orengo, C. A.; Moss, D. S.; Thornton, J. M. *Protein Sci.* **1993**, *2*, 1811.
(68) Chou, P. Y.; Fasman, G. D. *Annu. Rev. Biochem.* **1978**, *47*, 251.
(69) Creamer, T. P.; Srinivasan, R.; Rose, G. D. *Biochemistry* **1995**, *34*, 16245.
(70) Colloc'h, N.; Etchebest, C.; Thoreau, E.; Henrissat, B.; Mornon, J.-P. *Protein Eng.* **1993**, *6*, 377.
(71) Frishman, D.; Argos, P. *Proteins: Struct., Funct., Genet.* **1995**, *23*, 566.
(72) Kabsch, W.; Sander, C. *Biopolymers* **1983**, *22*, 2577.
(73) Murphy, K. P.; Bhakuni, V.; Xie, D.; Freire, E. *J. Mol. Biol.* **1992**, *227*, 293.
(74) Gill, S. J.; Wadsö, I. *Proc. Natl. Acad. Sci. U.S.A.* **1976**, *73*, 2955.
(75) Makhatadze, G. I.; Privalov, P. L. *J. Mol. Biol.* **1990**, *213*, 375.
(76) Gómez, J.; Hilser, V. J.; Xie, D.; Freire, E. *Proteins* **1995**, *22*, 404.
(77) Myers, J. K.; Pace, C. N.; Scholtz, J. M. *Protein. Sci.* **1995**, *4*, 2138.
(78) Graziano, G.; Barone, G. *J. Am. Chem. Soc.* **1996**, *118*, 1831.
(79) Murphy, K. P.; Gill, S. J. *Thermochim. Acta* **1990**, *172*, 11.
(80) Privalov, P. L.; Khechinashvili, N. N. *J. Mol. Biol.* **1974**, *86*, 665.
(81) Doig, A. J.; Williams, D. H. *Biochemistry* **1992**, *31*, 9371.
(82) Lee, B. K. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 5154.
(83) Baldwin, R. L.; Muller, N. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 7110.
(84) Yang, A.-S.; Sharp, K. A.; Honig, B. *J. Mol. Biol.* **1992**, *227*, 889.
(85) Murphy, K. P. *Biophys. Chem.* **1994**, *51*, 311.
(86) Barone, G.; Del Vecchio, P.; Giancola, C.; Graziano, G. *Int. J. Biol. Macromol.* **1995**, *17*, 251.
(87) Hilser, V. J.; Gómez, J.; Freire, E. *Proteins* **1996**, *26*, 123.
(88) Xie, D.; Freire, E. *Proteins* **1994**, *19*, 291.
(89) D'Aquino, J. A.; Gómez, J.; Hilser, V. J.; Lee, K. H.; Amzel, L. M.; Freire, E. *Proteins* **1996**, *25*, 143.
(90) Makhatadze, G. I.; Clore, G. M.; Gronenborn, A. M.; Privalov, P. L. *Biochemistry* **1994**, *33*, 9327.
(91) Hilser, V. J.; Freire, E. *J. Mol. Biol.* **1996**, *262*, 756.
(92) Dill, K. A. *J. Biol. Chem.* **1997**, *272*, 701.
(93) Mark, A. E.; van Gunsteren, W. F. *J. Mol. Biol.* **1994**, *240*, 167.
(94) Shrake, A.; Rupley, J. A. *J. Mol. Biol.* **1973**, *79*, 351.
(95) Pace, C. N.; Laurents, D. V.; Thomson, J. A. *Biochemistry* **1990**, *29*, 2564.
(96) Evans, P. A.; Topping, K. D.; Woolfson, D. N.; Dobson, C. M. *Proteins: Struct., Funct., Genet.* **1991**, *9*, 248.
(97) Sosnick, T. R.; Trewhella, J. *Biochemistry* **1992**, *31*, 8329.
(98) Neri, D.; Billeter, M.; Wider, G.; Wüthrich, K. *Science* **1992**, *257*, 1559.
(99) Fink, A. L.; Calciano, L. J.; Goto, Y.; Kurotsu, T.; Palleros, D. R. *Biochemistry* **1994**, *33*, 12504.
(100) Pace, C. N.; Laurents, D. V.; Erickson, R. E. *Biochemistry* **1992**, *31*, 2728.
(101) Brandts, J. F. *J. Am.. Chem. Soc.* **1964**, *86*, 4302.
(102) Stites, W. E. *Chem. Rev.* **1997**, *97*, 1233 (accompanying article in this issue).

CR960383C