PROTEINS: Structure, Function, and Genetics 22:404–412 (1995)

The Heat Capacity of Proteins

Javier Gómez, Vincent J. Hilser, Dong Xie, and Ernesto Freire
Department of Biology and the Biocalorimetry Center, The Johns Hopkins University, Baltimore, Maryland 21218

ABSTRACT **The heat capacity plays a major role in the determination of the energetics of protein folding and molecular recognition. As such, a better understanding of this thermodynamic parameter and its structural origin will provide new insights for the development of better molecular design strategies. In this paper we have analyzed the absolute heat capacity of proteins in different conformations. The results of these studies indicate that three major terms account for the absolute heat capacity of a protein: (1) one term that depends only on the primary or covalent structure of a protein and contains contributions from vibrational frequencies arising from the stretching and bending modes of each valence bond and internal rotations; (2) a term that contains the contributions of noncovalent interactions arising from secondary and tertiary structure; and (3) a term that contains the contributions of hydration. For a typical globular protein in solution the bulk of the heat capacity at 25°C is given by the covalent structure term (close to 85% of the total). The hydration term contributes about 15 and 40% to the total heat capacity of the native and unfolded states, respectively. The contribution of non-covalent structure to the total heat capacity of the native state is positive but very small and does not amount to more than 3% at 25°C. The change in heat capacity upon unfolding is primarily given by the increase in the hydration term (about 95%) and to a much lesser extent by the loss of noncovalent interactions (up to ~5%). It is demonstrated that a single universal mathematical function can be used to represent the partial molar heat capacity of the native and unfolded states of proteins in solution. This function can be experimentally written in terms of the molecular weight, the polar and apolar solvent accessible surface areas, and the total area buried from the solvent. This unique function accurately predicts the different magnitude and temperature dependences of the heat capacity of both the native and unfolded states, and therefore of the heat capacity changes associated with folding/unfolding transitions.** © 1995 Wiley-Liss, Inc.

Key words: protein thermodynamics, protein folding, protein stability, protein thermodynamics, energetics, protein design

INTRODUCTION

The heat capacity function occupies a central role in the determination of the energetics of stabilization of protein structures. It has been known for many years that the unfolding of a protein is accompanied by a positive heat capacity change and that this change greatly influences the behavior of the Gibbs energy of stabilization.[1] For a typical globular protein of ~15 kDa the change in heat capacity is on the order of 1.5–2.5 kcal/K·mol and has a profound effect on the overall energetics. The heat capacity change, ΔC_p, affects both the enthalpy, $\Delta H(T)$, and entropy, $\Delta S(T)$, changes that determine the Gibbs energy,

$$\Delta H(T) = \Delta H(T_R) + \int_{T_R}^{T} \Delta C_p \, dT \qquad (1a)$$

$$\Delta S(T) = \Delta S(T_R) + \int_{T_R}^{T} \Delta C_p \, d \ln T \qquad (1b)$$

where $\Delta H(T_R)$ and $\Delta S(T_R)$ are the enthalpy and entropy changes at the reference temperature (T_R). For small globular proteins the second term in Eq. (1a) modifies the enthalpy of stabilization by 150–250 kcal/mol within the temperature interval of 0–100°C. Similarly, the entropy change exhibits a variation of 450–780 cal/K·mol within the same temperature interval. This change in heat capacity is responsible for the existence of phenomena like cold denaturation or the presence of a temperature at which proteins exhibit a maximal stability. Also, the existence of a ΔC_p adds an additional complication to structure-based energetic computations. It is apparent that an accurate and rigorous understanding of protein energetics requires an equally accurate and rigorous understanding of the heat capacity function.

It was realized many years ago that the positive change in heat capacity that accompanies unfolding is mainly due to the exposure to the solvent of apolar groups that are buried in the native state.[1] Most recently, it has been realized that the hydration of both polar and apolar groups contribute to the heat capacity change, albeit with opposite signs.[2, 3] During the last few years our laboratory and others

Received January 6, 1995; revision accepted March 7, 1995.
Address reprint requests to Ernesto Freire, Department of Biology and the Biocalorimetry Center, The Johns Hopkins University, Baltimore, MD 21218.

have shown that the change in heat capacity associated with protein folding/unfolding can be parameterized in terms of changes in solvent accessible apolar and polar surface areas and that this approach provides an accurate representation of the heat capacity changes.[3–6] The results of those laboratories provide the strongest evidence that protein energetics cannot be understood without explicitly considering the role of the solvent.

Recently, it has become evident that the change in heat capacity upon folding/unfolding is not independent of temperature as supposed earlier.[1, 7] In fact, it progressively diminishes at higher temperatures and in some cases also at low temperatures.[7] This effect is due to the fact that the heat capacity of the native state increases more or less linearly and with a faster temperature rate than that of the unfolded state. In addition, the heat capacity of the unfolded state exhibits a decreasing slope at higher temperatures. A successful structural parameterization of the heat capacity change and its temperature dependence requires an explicit account of the different behavior of the various conformational states of a protein. Until today, however, it has not been possible to account for the heat capacity of different protein conformations, including the native and unfolded states, within a unified framework. The purpose of this paper is to present an initial attempt to model the heat capacity of different states with a single mathematical function.

The partial molar heat capacity and its temperature dependence have been measured calorimetrically for the native and unfolded states of several proteins (hen egg white lysozyme, ribonuclease A, myoglobin, cytochrome *c*, barnase, BPTI, interleukin 1β, staphylococcal nuclease, ribonuclease T1, ubiquitin, and α-lactalbumin).[7–12] Since high resolution structural information is also available for most of these proteins, they provide a unique opportunity to explore the structural basis of its heat capacity.

THE HEAT CAPACITY AT CONSTANT PRESSURE

The heat capacity at constant pressure, C_p, measures directly the magnitude of the enthalpy fluctuations in a system. From the point of view of statistical thermodynamics, the mean enthalpy of a system, $<H>$, is given by the equation

$$\langle H\rangle = \sum_{i=0}^{N} H_i \cdot \exp(-G_i/RT)/Q \qquad (2)$$

where the canonical partition function, Q, is equal to

$$Q = \sum_{i=0}^{N} \exp(-G_i/RT)$$

$G_i = H_i - T{\cdot}S_i$ is the Gibbs energy and R the gas constant. The heat capacity, C_p, is the temperature derivative of Eq. (2) obtained at constant pressure,

$$C_p = (\partial\langle H\rangle/\partial T)_p = \left\{\left[\sum_{i=0}^{N} H_i^2 \cdot \exp(-G_i/RT)/Q\right] - \left[\sum_{i=0}^{N} H_i \cdot \exp(-G_i/RT)/Q\right]^2\right\}/RT^2$$

$$= \{\langle H^2\rangle - \langle H\rangle^2\}/RT^2 \qquad (3b)$$

Equation (3) is a standard result in statistical thermodynamics and indicates that the heat capacity at constant pressure is directly proportional to the second moment of the enthalpy distribution about its mean or more simply to the dispersion of the enthalpy distribution. Theoretical studies of the heat capacity of proteins suggest that fluctuations in enthalpy arise primarily from hindered internal rotations (e.g., methyl groups), low frequency conformational fluctuations, and high frequency bond stretching and bending modes.[13, 14] Other effects like noncovalent interactions and the hydration of groups located at the protein surface are expected to contribute a smaller fraction of the total heat capacity of a protein, even though they are expected to account for most of the change in heat capacity associated with a change in protein conformation.

THE HEAT CAPACITY OF PROTEINS

Experimentally, the partial molar heat capacity, C_p, of proteins in aqueous solution has been considered as being composed of an intrinsic term and a term due to hydration (see, for example, refs. 15, 16). The intrinsic term can be further decomposed into a term containing the contributions from covalent bonds and another term arising from noncovalent interactions. Experimental studies on a large number of organic molecules indicate that the heat capacity of those compounds is largely additive at the bond or group level suggesting that only local effects contribute significantly to the covalent contribution to the heat capacity (see, for example, ref. 13). The contribution of hydration to the heat capacity appears to be complete when a single water monolayer covers the protein.[16] If this is the case, the heat capacity of a protein can be written as

$$C_p = C_{p,a} + C_{p,b} + C_{p,c} + C_{p,p} \qquad (4)$$

The first term, $C_{p,a}$, will be called the primary heat capacity and contains the atomic and covalent bond contributions. By definition, this term depends only on the amino acid composition of the protein and is independent of its conformational state. The second term, $C_{p,b}$, contains the contributions of all noncovalent interactions within the protein molecule. The third term, $C_{p,c}$, contains the contributions due to the interactions of the protein with the solvent, i.e., hydration. $C_{p,p}$ is the heat capacity due to the state

TABLE I. Experimental and Calculated Heat Capacities of Individual Amino Acids

Amino acid	C_p* (cal/K·mol)	C_p† (cal/K·mol)	C_p‡ (cal/K·mol)	Amino acid	C_p* (cal/K·mol)	C_p† (cal/K·mol)	C_p‡ (cal/K·mol)
Ala	29.22	27.13	26.94	Leu	48.03	43.75	43.32
Arg	55.8	55.65	56.02	Lys	48.94	45.1	48.28
Asn	38.3	38.8	38.4	Met	69.32	43.0	42.68
Asp	37.09	37.1	36.5	Phe	48.52	53.3	44.7
Cys	38.8	32.1	31.8	Pro	36.13	37.45	35.96
Gln	44.02	44.25	43.86	Ser	32.4	30.8	30.6
Glu	41.84	42.55	41.96	Thr	35.2	32.5	36.06
Gly	23.71	21.95	21.48	Trp	56.92	65.05	55.98
His	51.48	48.0	51.58	Tyr	51.73	56.7	51.9
Ile	45.0	43.75	43.32	Val	40.35	38.3	37.86

*Experimental literature values at 25°C from Hutchens et al.[21].

† Values calculated assuming atomic additivity parameters tabulated by Benson.[20] The atomic values used in the calculations are H = 0.85, C = 3.75, N = 3.4, O = 3.4, S = 4.7 cal/K·mol respectively.

‡Values calculated assuming bond additivity parameters tabulated by Benson.[20] the bond values used in the calculation of each amino acid contribution are C–H = 1.74, C–C = 1.98, C–N = 2.1, C–C(O) = 3.7, C(O)–O = 2.2, C(O)–N = 2.2, Ph–C = 4.5, Ph–O = 4.5, Ph–H = 3.0, Ph–N = 4.5, =C–C = 2.6, =C–H = 2.6, =C–N = 2.6, C–S = 3.4, C–O = 2.7, S–H = 3.2, N–H = 2.3, and O–H = 2.7 cal/K·mol, respectively. See also the work by Janz.[17]

of protonation of those side chains with ionizable groups (e.g., His, Asp, Glu, Arg, Lys). $C_{p,b}$, $C_{p,c}$ and $C_{p,p}$ depend on the physical state of the protein, the secondary structure content, and interactions with the solvent, respectively.

According to the definitions above, $C_{p,a}$ would experimentally correspond to the heat capacity of the unfolded peptide chain in the anhydrous state; the sum $C_{p,a} + C_{p,b}$ would be equal to the heat capacity of the anhydrous protein in the native state. The change in heat capacity associated with folding/unfolding or conformational transitions and also with protein/protein associations in solution involves only the terms $C_{p,b}$, $C_{p,c}$, and $C_{p,p}$ since these protein transformations do not involve changes in mass or primary structure,

$$\Delta C_p = \Delta C_{p,b} + \Delta C_{p,c} + \Delta C_{p,p} \quad (5)$$

THE HEAT CAPACITY OF ANHYDROUS PROTEINS

Within the temperature range of interest in biology (~0–100°C) the primary heat capacity of a protein predominantly contains contributions from vibrational frequencies arising from the stretching and bending modes of each valence bond and internal rotations.[17] Electronic contributions are negligible in this temperature range.[18] The heat capacity of all 20 amino acids has been measured in the anhydrous state as well as that of some dipeptides.[19] In addition, individual atomic and bond contributions have been tabulated from the analysis of the heat capacities of small organic compounds.[17, 18, 20] These experimental values can be used to estimate the magnitude of the primary heat capacity of a protein.

Table I summarizes the experimental and calculated heat capacities of all 20 amino acids in the anhydrous state at 25°C as well as the corresponding values calculated by using the contributions for individual atomic and bond contributions tabulated by Benson.[13] In general, the two additivity schemes underestimate slightly the experimental heat capacity values. Table II summarizes experimental values for those proteins for which data in the anhydrous state are available in the literature. Also shown in Table II are the calculated values for the primary heat capacity of those proteins. As shown in Table II, the heat capacity of the anhydrous protein is very close in magnitude to the one calculated from the contributions of individual amino acids plus the additional contribution of the peptide bond, consistent with the idea that the bulk of the heat capacity of a protein originates from the covalent structure. It should be noticed, however, that the experimental values for the anhydrous native protein are generally slightly larger than the calculated primary values suggesting that noncovalent interactions do contribute, albeit slightly to the heat capacity. On the average, the contribution of the primary structure to the heat capacity of the anhydrous protein accounts for 91–96% of the total.

The heat capacity of anhydrous proteins increases linearly with temperature within the temperature range of interest.[19, 21] Also, the heat capacity of anhydrous proteins when normalized on a weight basis is very similar for all proteins as indicated in Table II and averages 0.298±0.003 cal/K·g. The same situation occurs for the temperature dependence, which averages $9.77 \pm 0.2 \times 10^{-4}$ cal/K^2·g for all proteins. Thus, on the average the primary heat capacity of a globular protein appears to be directly proportional to its mass or molecular weight. Certainly the same proportionality constant will not hold for small peptides or peptides in which the amino acid composition largely differs from that of a globular protein. If the difference between the reported values for the heat capacity of the anhydrous native proteins and the

TABLE II. Heat Capacities of Anhydrous Proteins

Protein	$C_{p,exp}(25)$ (cal/K·g)	$C_p(25)$* (cal/K·g)	$C_p(25)^{\dagger}$ (cal/K·g)	$C_p(25)^{\ddagger}$ (cal/K·g)	$\partial C_{p,exp}/\partial T$ (cal/K^2·g)
Albumin[§]	0.305	0.291	0.274	0.2683	9.89×10^{-4}
Albumin**	0.295				—
Chymotrypsinogen[§]	0.309	0.282	0.268	0.265	9.92×10^{-4}
Chymotrypsinogen**	0.293				—
Insulin[§]	0.300	0.283	0.271	0.264	9.5×10^{-4}
Lysozyme**	0.285	0.282	0.270	0.264	—
Lysozyme[††]	0.301				

*Calculated primary heat capacity using experimental heat capacities for anhydrous amino acids. $C_p = \Sigma C_p$ (amino acid) + $(N-1)\cdot C_p$(peptide bond formation), where C_p (peptide bond formation) is the effective average contribution due to the formation of a peptide bond and the release of a water molecule. In the solid state this value amounts to -9.33 ± 0.74 cal/K·mol and was calculated as the averaged difference between the experimental heat capacities of some anhydrous dipeptides (Gly-Gly, Ala-Gly, and Leu-Gly) and the heat capacities of the individual anhydrous amino acids.[15,21]

†Calculated primary heat capacity using the amino acid heat capacities calculated from the additivity of atomic values tabulated by Benson.[20]

‡Calculated primary heat capacity using the amino acid heat capacities calculated from the additivity of bond values tabulated by Benson.[20]

§Hutchens et al.[21]

**Suurkuusk.[15]

††Yang and Rupley.[16]

values obtained by adding the contributions of the amino acids is taken as an indication of the contributions of secondary structure, then it follows that about 95% of the total heat capacity of the anhydrous protein (~0.28 cal/K·g) is given by the primary heat capacity.

THE HEAT CAPACITY OF NATIVE PROTEINS IN SOLUTION

The heat capacity of native proteins in solution exhibits an almost linear temperature dependence in the temperature interval in which it can be measured (usually 0–50°C since the onset of thermal denaturation precludes measurement at higher temperatures). Table III summarizes the heat capacities at 25°C and the corresponding temperature dependences for the proteins in the database. In all cases the regression coefficient is better than 0.99 for the linear least squares analysis.

Several conclusions can be obtained from an inspection of the data in Table III. First, it is evident that at 25°C a large fraction of the heat capacity is given by the primary heat capacity. About 85% of the total heat capacity of the native state in solution is given by the covalent structure. Second, at 25°C the heat capacity of the native proteins in solution is larger than that of the anhydrous proteins indicating that, on the whole, hydration contributes positively to the heat capacity of the native state. The relative increase, however, is not the same for all proteins suggesting that the composition of the protein surface mediates the magnitude of this increase. Also, at 25°C the temperature dependence of the heat capacity of the native protein in solution is larger than that of the anhydrous protein and larger than that of the unfolded polypeptide (see below).

In solution, the additional contribution to the heat capacity is given by the solvation of those atoms located at the protein surface. The solvent-exposed surface of native proteins is composed of polar and apolar regions in different proportions. On the average, about 55% of the total solvent accessible surface area in the native state is apolar, which qualitatively explains the positive contribution of hydration. It is expected that the hydration heat capacity of a protein should be proportional to the dimensions of the apolar and polar solvent accessible surface areas in much the same way as the heat capacity changes associated with conformational transitions.[4, 5]

As mentioned above, the difference in heat capacity between the anhydrous native protein and the primary heat capacity calculated from the contributions of individual amino acids is very small suggesting that the contribution of noncovalent interactions is also small. This conclusion is in agreement with previous results (see, for example, Eftink et al.[22]). Noncovalent interactions are expected to be a function of the packing density within the protein and, as a first approximation, are expected to scale in terms of the total buried surface area of the protein. Since the contribution of this term to the total heat capacity is very small (Fig. 1) a more detailed analysis that explicitly includes variation in packing densities is not possible at the present time. Figure 1 illustrates the magnitude of the different contributions to the heat capacity of lysozyme, the only protein for which all the individual values have been measured experimentally.

THE HEAT CAPACITY OF THE UNFOLDED STATE AND THE CONTRIBUTION OF HYDRATION

While the heat capacity of the native state has a linear temperature dependence, the heat capacity of

TABLE III. Heat Capacities of Native Proteins in Solution*

Protein	$C_p(25)$ (cal/K·mol)	$C_p(25)$ (cal/K·g)	$\partial C_p/\partial T$ (cal/K^2·mol)	$\partial C_p/\partial T$ (cal/K^2·g)
Cytochrome c[a]	4192.9	0.3276	20.744	1.62×10^{-3}
Lysozyme[a]	4780.2	0.3341	21.334	1.49×10^{-3}
Myoglobin[a]	5794.5	0.3255	29.302	1.65×10^{-3}
Staph. nuclease[b]	5410.0	0.338	26.000	1.63×10^{-3}
RNase A[a]	4982.3	0.363	15.959	1.17×10^{-3}
BPTI[c]	2270.6	0.348	9.3140	1.43×10^{-3}
Barnase[d]	4425.9	0.359	23.923	1.94×10^{-3}
Interleukin 1β[e]	6713.3	0.386	19.899	1.14×10^{-3}
RNase T1[f]	3851.7	0.348	17.3	1.56×10^{-3}
Ovalbumin[g]	15444	0.351	81.4	1.85×10^{-3}
Chymotrypsinogen[g]	8854.8	0.345	44.9	1.75×10^{-3}
Ubiquitin[h]	3014.4	0.353	16.322	$1.91 \times 10^{\Sigma 3}$

*In this and other tables and figures, data for the referenced proteins were obtained from the following papers: (a) Privalov et al.[7] (b) Xie et al.[12] (c) Makhatadze et al.[9] (d) Griko et al.[8] (e) Makhatadze et al.[10] (f) Yu et al.[26] (g) Unpublished results from this laboratory. (h) Wintrode et al.[27]

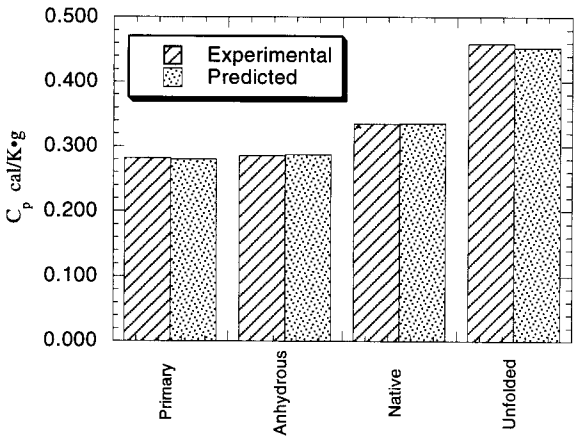


Fig. 1. The different contributions to the heat capacity of hen egg white lysozyme at 25°C. While the bulk of the heat capacity is given by the contributions from the primary structure, the major contribution to the heat capacity change upon unfolding is given by the hydration term. The anhydrous heat capacity is from Suurkuusk,[15] the contribution of primary structure was calculated from the experimental amino acid values,[21] and the values in solution are from Privalov and Makhatadze.[7] The predicted values were calculated with the parameters in Table V. In this paper, all the surface areas calculations were performed as described before[4] using the implementation of Lee and Richard's algorithm[28] in the program ACCESS (Scott R. Presnell, University of California, San Francisco), with a probe radius of 1.4 Å and a slice width of 0.25 Å.

the unfolded state does not. Within the temperature interval 0–100°C the heat capacity of the unfolded state is well approximated by a second order polynomial on temperature.[11, 12, 23] For convenience our series expansion will be centered around 25°C:

$$C_p = C_p(25) + a\cdot(T - 25) + b\cdot(T - 25)^2 \quad (6)$$

Table IV summarizes the results of the second order polynomial fit for each of the proteins in the database. As shown in table IV, for all proteins a second-order polynomial accurately represents the heat capacity of the unfolded state with a regression coefficient better than 0.995 for the entire temperature interval covered by the data (5–125°C). On average, at 25°C the heat capacity of the unfolded state is about 1.45 times larger than that of the native state, whereas at 100°C it is only 1.15 larger. This difference in heat capacity and its temperature dependence can be accounted for in terms of the hydration of those protein groups that are buried from the solvent in the native state and become exposed to the solvent in the unfolded polypeptide. In general, the unfolded state has a higher proportion of apolar residues than the native state and this difference is expected to contribute significantly to the different temperature dependences of their heat capacities. To a lesser extent, the difference in heat capacity must be also proportional to the loss of noncovalent interactions (secondary and tertiary structure) reflected in $C_{p,b}$. It has been shown before that the heat capacity due to hydration can be well approximated in terms of solvent accessible polar (ASA_{ap}) and polar (ASA_{pol}) surfaces. If this is the case, it should be possible to express the hydration heat capacity using a unique function of ASA_{ap} and ASA_{pol} for the native and unfolded states. More precisely, it should be possible to express the hydration heat capacity of any arbitrary state of a protein as a quadratic function of temperature in which each coefficient is a linear combination of ASA_{ap} and ASA_{pol}.

IONIZATION CONTRIBUTIONS TO THE HEAT CAPACITY FUNCTION

Side chains with ionizable groups (e.g., His, Asp, Glu, Arg, Lys) contribute differently to the heat capacity depending on whether they are protonated or not. These contributions are, however, very small in relation to the overall magnitude of the heat capacity. For example, the heat capacity of a protonated imidazol group is 4 cal/K·mol larger than that of an unprotonated one; and that of a protonated carbox-

TABLE IVa. Heat Capacities of Unfolded Proteins in Solution: $C_p = C_p(25) + a\cdot(T - 25) + b\cdot(T - 25)^2$

Protein	$C_p(25)$ (cal/K·mol)	a (cal/K^2·mol)	b (cal/K^3·mol)	R
Cytochrome c	5799	15.50	−0.0953	0.997
Lysozyme	6568	23.84	−0.143	0.999
Myoglobin	8294	28.61	−0.173	0.996
RNase A	6200	18.50	−0.101	0.997
BPTI	2958	8.29	−0.052	0.996
Barnase	6041	16.19	−0.096	0.995
Interleukin 1β	8484	21.75	−0.130	0.997
RNase T1	5130	15.43	−0.065	0.999
Ubiquitin	4378	11.996	−0.0737	0.996

TABLE IVb. Heat Capacities of Unfolded Proteins in Solution: $C_p = C_p(25) + a\cdot(T - 25) + b\cdot(T - 25)^2$

Protein	$C_p(25)$ (cal/K·g)	a (cal/k^2·g)	b (cal/K^3·g)	R
Cytochrome c	0.453	1.21×10^{-3}	-7.44×10^{-6}	0.997
Lysozyme	0.459	1.67×10^{-3}	-9.99×10^{-6}	0.999
Myoglobin	0.466	1.61×10^{-3}	-9.72×10^{-6}	0.996
RNase A	0.453	1.35×10^{-3}	-7.38×10^{-6}	0.997
BPTI	0.454	1.27×10^{-3}	-7.98×10^{-6}	0.996
Barnase	0.490	1.31×10^{-3}	-7.797×10^{-6}	0.995
Interleukin 1β	0.488	1.25×10^{-3}	-7.471×10^{-6}	0.997
RNase T1	0.4635	1.39×10^{-3}	-5.873×10^{-6}	0.999
Ubiquitin	0.513	1.41×10^{-3}	-8.636×10^{-6}	0.996

ylic group is about 30 cal/K·mol larger than that of an unprotonated one. This term is directly proportional to the degree of protonation, F_p, and can be written as $F_p\cdot\Delta C_{p,p}$, where $\Delta C_{p,p}$ is the protonation heat capacity. In many cases the degree of protonation changes upon unfolding, in which case this term contributes $\Delta F_p\cdot\Delta C_{p,p}$ to the overall ΔC_p. There is still another contribution from protonation to the heat capacity, a term that arises from thermal fluctuations in the degree of protonation. This term is proportional to the effective enthalpy of protonation, ΔH_p, and is given by $F_p\cdot(1 - F_p)\cdot\Delta H_p^2/R\cdot T^2$. This contribution is maximal when $F_p = 0.5$, i.e., when the pH is equal to the pK_a of the ionizable group. For example, for a histidine in a nonbuffered solution at pH = pK_a the maximal contribution due to thermal fluctuations is expected to be around 70 cal/K·mol, since under these conditions $\Delta H_p = -7$ kcal/mol. Under usual conditions, however, the total contribution due to protonation is expected to be small, especially since both the heat capacity and enthalpy changes associated with the release or absorption of protons are generally opposed by the accompanying heat capacity and enthalpy changes in the buffer.

GLOBAL FIT OF THE HEAT CAPACITY

The discussion presented above provides some constraints to the functional form of the heat capacity function. In the absence of or after correction for protonation effects, it appears reasonable to express the heat capacity of any protein conformation by Eq. (4) ($C_p = C_{p,a} + C_{p,b} + C_{p,c}$) and that this equation can be used to perform a global nonlinear least squares fit of the entire database independently of the conformational state of the proteins. In other words, the same equation with the same parameters should be able to predict the linear temperature dependence of the heat capacity of the native state and the nonlinear temperature dependence of the heat capacity of the unfolded state. As discussed above, $C_{p,a}$ scales in terms of the molecular weight of a protein while $C_{p,b}$ and $C_{p,c}$ should be proportional to both, the areas of the protein that are buried from the solvent and those that are exposed to the solvent. If this is the case, it should be possible to accurately express the heat capacity of any protein state if the molecular weight, MW, the total area buried from the solvent, BSA_{Total}, and the solvent accessible apolar, ASA_{ap}, and polar, ASA_{pol}, surface areas are known. According to our discussion, each term in Eq. (4) is given by

$$C_{p,a} = [v + w\cdot(T - 25)]\cdot MW \quad (7)$$

$$C_{p,b} = [p + q\cdot(T - 25)]\cdot BSA_{Total} \quad (8)$$

$$C_{p,c} = a(T)\cdot ASA_{app} + b(T)\cdot ASA_{pol} \quad (9)$$

where

$$a(T) = a_1 + a_2\cdot(T - 25) + a_3\cdot(T - 25)^2 \quad (10)$$

$$b(T) = b_1 + b_2\cdot(T - 25) + b_3\cdot(T - 25)^2 \quad (11)$$

Equations (7)–(11) can be used to fit the entire database irrespective of protein conformation in order to obtain the best values for the model parameters.

TABLE V. Global Fitting Parameters for Heat Capacity Function

v (cal/K·g)	0.28
w (cal/K^2·g)	9.75×10^{-4}
p (cal/K·mol Å^2)	8.7×10^{-3}
q (cal/K^2·molÅ^2)	6.43×10^{-4}
a_1 (cal/K·mol Å^2)	0.45
a_2 (cal/K^2·molÅ^2)	2.63×10^{-4}
a_3 (cal/K^3·molÅ^2)	-4.2×10^{-5}
b_1 (cal/K·mol Å^2)	−0.265
b_2 (cal/K^2·molÅ^2)	2.85×10^{-4}
b_3 (cal/K^3·molÅ^2)	4.31×10^{-5}

The best set of parameters is the one that minimizes the sum of squared residuals [SSR = Σ ($C_{p,calc}$ − $C_{p,exp}$)^2]. Table V summarizes the results of the global non-linear least squares fit of the database. Figure 2A–I shows the experimental and calculated heat capacities for the native and unfolded states of the proteins in the database. As seen in Figure 2 the parameterization provided by Eqs. (7)–(11) reproduces the experimental values within the experimental limits throughout the entire temperature interval (5–125°C). In particular, a single mathematical function and a unique set of parameters predict the almost linear temperature dependence of the heat capacity of the native state and the progressively decreasing temperature dependence of the unfolded state. The only case in which a significant deviation is observed is for the native state of interleukin 1β; however, the reported specific heat capacity of this protein is also significantly higher than that found for other proteins.[10]

According to the results of the nonlinear least squares fit, the primary heat capacity at 25°C is equal to 0.28 cal/K·g, which is very close to the values calculated from individual amino acids and shown in Table II. At the same temperature the contribution of noncovalent interactions is on the order of 8.7×10^{-3} cal/K·mol Å^2 of total buried surface area. For a typical globular protein this value amounts to about 0.007 cal/K·g and indicates that noncovalent interactions contribute less than 3% of the total heat capacity of the anhydrous protein. The results for lysozyme are plotted in Figure 1 and graphically illustrate the relative magnitude of each term and the accuracy of the predicted values. As shown in Figure 1, at 25°C the hydration of the native state increases its heat capacity by 0.05 cal/K·g to about 0.34 cal/K·g. In solution, approximately 15% of the total heat capacity is given by the hydration contribution. For the unfolded state on the other hand, the heat capacity amounts to 0.46 cal/K·g indicating that under these conditions approximately 40% of the total heat capacity is given by the hydration contribution.

THE HEAT CAPACITY CHANGE UPON UNFOLDING

Two different effects are the primary contributors to the heat capacity change upon unfolding, the hydration of polar and apolar groups that are buried from the solvent, and the disruption of noncovalent interactions existing in the native state. Figure 3 illustrates the calculated temperature dependence of the three different contributions to ΔC_p. The heat capacity of apolar hydration is positive, however it decreases from 0.45 cal/K·mol Å^2 at 25°C to about 0.23 cal/K·mol Å^2 at 100°C. The heat capacity of polar hydration is negative. It amounts to −0.26 cal/K·mol Å^2 at 25°C but it becomes negligibly small around 100°C. The values obtained at 25°C are similar to those determined before using a temperature independent parameterization.[4,6] Noncovalent interactions contribute positively to the partial molar heat capacity of the native state; therefore their disruption upon unfolding contribute negatively to ΔC_p. This contribution is very small at 25°C but increases to about 0.06 cal/K·mol buried Å^2 at 100°C. The current results include the previous ones as a subset and extend the range of validity of the structural parameterization over the temperature interval 5–125°C. Also, the current results reconcile the structural parameterization with the results obtained by Privalov and co-workers[7] over the entire temperature range studied. The temperature dependence of the elementary contributions depicted in Figure 3 accounts quantitatively for the observed ΔC_p for unfolding and its decrease at high temperatures. The parameterization also predicts a slight decrease in ΔC_p at low temperatures. For the proteins in the database, the error is less than 10% over the entire temperature range. In general, the change in heat capacity due to hydration and noncovalent interactions can be expressed as

$$\Delta C_{p,c} = a(T)\cdot\Delta ASA_p + b(T)\cdot\Delta ASA_{pol} \quad (12)$$

$$\Delta C_{p,b} = [p + q\cdot(T - 25)]\cdot\Delta BSA_{Total} \quad (13)$$

CONCLUSIONS

The results of the analysis presented in this paper indicate that the absolute heat capacity of different protein conformations can be accurately calculated from structural parameters over a wide temperature range. It has been shown before that the heat capacity of the unfolded state can be accounted for in terms of the individual contributions of the amino acid side chains and the peptide backbone,[7] i.e., it exhibits group additivity. The heat capacities of the native state in solution or that of partly folded states on the other hand do not exhibit that type of additivity: they cannot be predicted from the amino acid sequence. They can be predicted, however, if the three-dimensional structure of the protein is known. The results presented in this paper indicate that, within the ex-

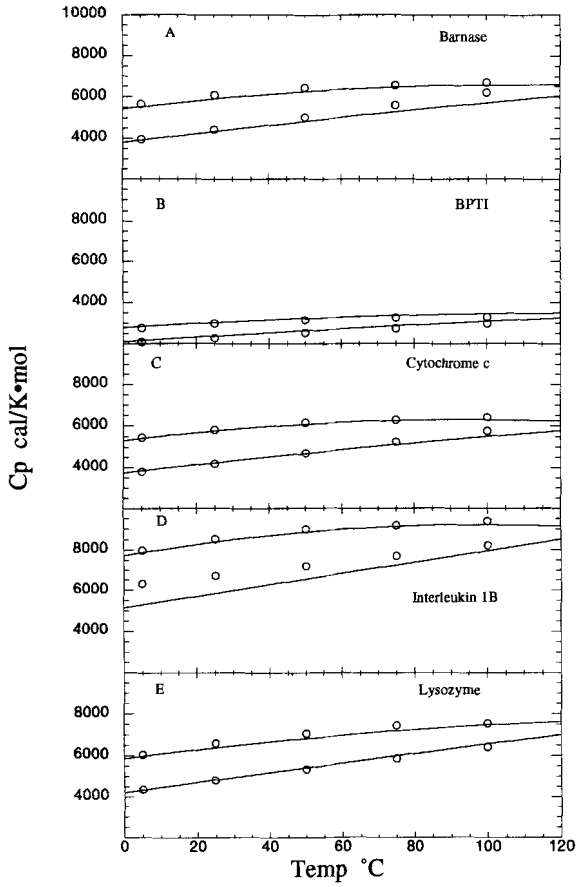
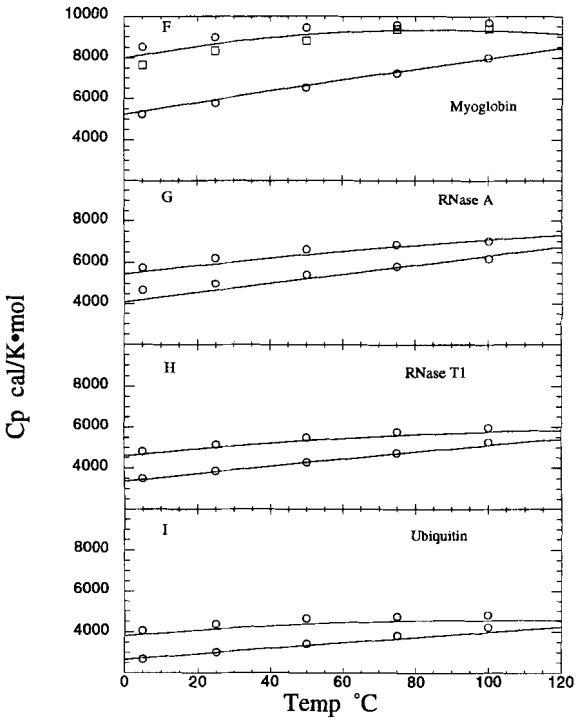


Fig. 2. Comparison between the experimental (circles) and calculated (solid lines) heat capacities for the native and unfolded states of nine proteins in the database (**A**, barnase; **B**, BPTI; **C**, cytochrome *c*; **D**, interleukin 1β; **E**, lysozyme; **F**, myoglobin; **G**, RNase A; **H**, RNase T1; **I**, ubiquitin). For myoglobin, the squares and the circles represent the experimental and the calculated values for the unfolded states obtained by Privalov and Makhatadze,[7] respectively. Data points for the native state above 50°C were extrapolated in the original references. The calculated values were obtained with Eq. (7)–(11) using the parameter values in Table V (see text for details).



perimental error, the heat capacity is additive in terms of the primary, noncovalent and hydration terms, which in turn can be expressed in terms of the molecular weight, the surface area buried from the solvent, and the exposed polar and apolar surfaces accessible to the solvent. A very important conclusion can be deduced from this observation. From a rigorous thermodynamic standpoint, if the heat capacity is additive on a set of system parameters then the enthalpy and entropy are also additive on those same parameters plus the addition of a constant term. For example, Eq. (1) becomes

$$\Delta H(T) = \Delta H(T_{R,H}) + \int_{T_{R,H}}^{T} \Delta C_{p,a}\, dT + \int_{T_{R,H}}^{T} \Delta C_{p,b}\, dT + \int_{T_{R,H}}^{T} \Delta C_{p,c}\, dT \qquad (12)$$

and similarly, for the entropy

$$\Delta S(T) = \Delta S(T_{R,H}) + \int_{T_{R,H}}^{T} \Delta C_{p,a}\, d \ln T + \int_{T_{R,H}}^{T} \Delta C_{p,b}\, d \ln T + \int_{T_{R,H}}^{T} \Delta C_{p,c}\, d \ln T \qquad (13)$$

Since the heat capacity is additive on the chosen set of system parameters within the entire temperature range of interest (0–100°C), it is clear that the enthalpy and entropy can be expressed accurately if appropriate reference temperatures are found at which these two quantities can be accurately parameterized. This is true even if the enthalpy or entropy, and hence the Gibbs energy, are not additive on those system parameters at the specified reference temperatures (see for example Mark and van Gunsteren[24]). Within the current approach, a reasonable

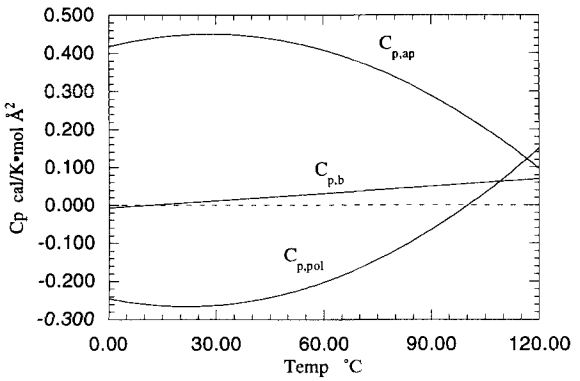


Fig. 3. Temperature dependence of elementary hydration contributions to the heat capacity ($C_{p,ap}$ and $C_{p,pol}$) per mol of Å^2 of polar and apolar protein surface exposed to the solvent. Also shown in the figure is the elementary contribution of noncovalent interactions ($C_{p,b}$) expressed per mol of Å^2 of buried protein surface.

strategy for structure based energetic predictions is to find the most appropriate reference temperatures for structural estimation of the constant terms $\Delta H(T_{R,H})$ and $\Delta S(T_{R,S})$.[6, 25] Another alternative is prediction of the temperatures at which the enthalpy and entropy changes are zero. If this could be done, then it would be possible to predict the Gibbs energy in terms of the heat capacity function only.

A necessary requirement for the development of successful strategies in structure-based molecular design is the ability to predict accurately the energetics of stabilization in the case of protein folding or the association energetics in the case of binding. The results presented in this paper represent a significant improvement over previous parameterization attempts that left the heat capacity as a constant independent of temperature. Also, the prediction of the heat capacity of the native state or that of partly folded conformations of proteins requires a structural parameterization in terms of elementary contributions rather than one based on group additivity since many times only portions of those groups are exposed to the solvent.

ACKNOWLEDGMENTS

We thank all the members of the Biocalorimetry Center for very helpful discussions. Supported by grants from the National Institutes of Health (RR-04328, GM-37911, and NS-24520) and the National Science Foundation (MCB-9118687).

REFERENCES

1. Privalov, P.L., Khechinashvili, N.N. A thermodynamic approach to the problem of stabilization of globular protein structure: A calorimetric study. J. Mol. Biol. 86:665–684, 1974.
2. Murphy, K.P., Gill, S.J. Solid model compounds and the thermodynamics of protein unfolding. J. Mol. Biol. 222: 699–709, 1991.
3. Privalov, P.L., Makhatadze, G.I. Contribution of hydration and noncovalent interactions to the heat capacity effect on protein unfolding. J. Mol. Biol. 224:715–723, 1992.
4. Murphy, K.P., Bhakuni, V., Xie, D., Freire, E. The molecular basis of co-operativity in protein folding III: Identification of cooperative folding units and folding intermediates. J. Mol. Biol. 227:293–306, 1992.
5. Spolar, R.S., Livingstone, J.R., Record, M.T., Jr. Use of liquid hydrocarbon and amide transfer data to estimate contributions to thermodynamic functions of protein folding from the removal of nonpolar and polar surface from water. Biochemistry 31:3947–3955, 1992.
6. Xie, D., Freire, E. Molecular basis of cooperativity in protein folding. V. Thermodynamic and structural conditions for the stabilization of compact denatured states. Proteins 19:291–301, 1994.
7. Privalov, P.L., Makhatadze, G.I. Heat capacity of proteins. II. Partial molar heat capacity of the unfolded polypeptide chain of proteins: Protein unfolding effects. J. Mol. Biol. 213:385–391, 1990.
8. Griko, Y., Makhatadze, G.I., Privalov, P.L., Hartley, R.W. Thermodynamics of barnase unfolding. Protein Sci. 3:669–676, 1994.
9. Makhatadze, G.I., Kim, K.-S., Woodward, C., Privalov, P.L. Thermodynamics of BPTI folding. Protein Sci. 2:2028–2036, 1993.
10. Makhatadze, G.I., Clore, G.M., Gronenborn, A., Privalov, P.L. Thermodynamics of unfolding of the all β-sheet protein Interleukin 1-β. Biochemistry 31:9327–9332, 1994.
11. Griko, Y., Freire, E., Privalov, P.L. Energetics of the alpha-lactalbumin states. A calorimetric and statistical thermodynamic study. Biochemistry 33:1889–1899, 1994.
12. Xie, D., Fox, R., Freire, E. Thermodynamic characterization of a Staphylococcal nuclease folding intermediate. Protein Sci. 3:2175–2184, 1994.
13. Benson, S.W. "Thermochemical Kinetics." New York: John Wiley, 1968.
14. Edelman, J. The low-temperature heat capacity of solid proteins. Biopolymers 32:209–218, 1992.
15. Suurkuusk, J. Specific heat measurements on lysozyme, chymotrypsinogen and ovalbumin in aqueous solution and in the solid state. Acta Chem. Scand. B 28:409–417, 1974.
16. Yang, P.H., Rupley, J.A. Protein-water interactions. Heat capacity of the lysozyme-water system. Biochemistry 18: 2654–2661, 1979.
17. Janz, G.J. "Estimation of Thermodynamic Properties of Organic Compounds." New York: Academic Press, 1958.
18. Blinder, S.M. On additivity rules in molecular thermodynamics. J. Am. Chem. Soc. 97:978–981, 1975.
19. Hutchens, J.O. In "Handbook of Chemistry and Selected Data for Molecular Biology." (eds. Sober, H.A., ed. Cleveland, OH: Chemical Rubber Co., 1970.
20. Benson, S.W., Buss, J.H. Additivity rules for the estimation of molecular properties: Thermodynamic properties. J. Chem. Phys. 29:546–572, 1958.
21. Hutchens, J.O., Cole, A.G., Stout, J.W. Heat capacities from 11 to 305 K and entropies of hydrated and anhydrous bovine zinc insulin and bovine chymotrypsinogen A. J. Biol. Chem. 224:26–32, 1969.
22. Eftink, M.R., Anusiem, A.C., Biltonen, R.L. Enthalpy-entropy compensation and heat capacity changes for protein-ligand interactions: General thermodynamic models and data for the binding of nucleotides to ribonuclease A. Biochemistry 22:3884–3896, 1983.
23. Martinez, J.C., Harrous, M.E., Filimonov, V.V., Mateo, P.L., Fersht, A.R. A calorimetric study of the thermal stability of barnase and its interaction with 3′GMP. Biochemistry 33:3919–3926, 1994.
24. Mark, A.E., van Gunsteren, W.F. Decomposition of the free energy of a system in terms of specific interactions. Implications for theoretical and experimental studies. J. Mol. Biol. 240:167–176, 1994.
25. Xie, D., Freire, E. Structure based prediction of protein folding intermediates. J. Mol. Biol. 242:62–80, 1994.
26. Yu, Y., Makhatadze, G.I., Pace, C.N., Privalov, P.L. Energetics of ribonuclease T1 structure. Biochemistry 33:3312–3319, 1994.
27. Wintrode, P.L., Makhatadze, G.I., Privalov, P.L. Thermodynamics of ubiquitin unfolding. Proteins 18:246–253, 1994.
28. Lee, B., Richards, F.M. The interpretation of protein structures: Estimation of static accessibility. J. Mol. Biol. 55: 379–400, 1971.