

Thermodynamic problems in structural molecular biology*

Peter L. Privalov‡

Department of Biology, Johns Hopkins University, Baltimore, MD 21218, USA

Abstract: The most essential feature of living biological systems is their high degree of structural organization. The key role is played by two linear heteropolymers, the proteins and nucleic acids. Under environmental conditions close to physiological, these biopolymers are folded into unique native conformations, genetically determined by the arrangement of their standard building blocks. In their native conformation, biological macromolecules recognize their partners and associate with them, forming specific, higher-order complexes, the “molecular machines”. Folding of biopolymers into their native conformation and their association with partners is in principle a reversible, thermodynamically driven process. Investigation of the thermodynamics of these basic biological processes has prime importance for understanding the mechanisms of forming these supra-macromolecular constructions and their functioning.

Keywords: microcalorimetry; macromolecules; DNA; proteins; hydration; thermodynamics.

INTRODUCTION

The phenomenon of life intrigues us all. It was not that long ago that we first dived in at its molecular level and found two remarkable molecular objects: the proteins and nucleic acids, both large linear polymers. The nucleic acids, particularly DNA, are assembled from four different standard blocks, the nucleotides, the sequence of which codes the genetic information. Proteins are assembled from 20 different amino acids arranged according to this genetic information. The most striking feature of these linear heteropolymers is their ability to fold into unique spatial conformations dictated by the arrangement of their building blocks. In the folded state, proteins recognize their partners, other proteins or nucleic acids, and form with them supra-macromolecular constructions—the so-called molecular machines. Some of these machines are involved in the utilization of the genetic information included in the DNA, others process various reactions and channel energy flows.

The folding of proteins and nucleic acids into their active, so-called “native” conformations and their subsequent association with specific partners are therefore two of the most basic biological functions. These processes are, however, interesting not only for biologists. They are also of special interest for thermochemists and particularly calorimetrists since these are reversible, thermodynamically driven processes of transforming linear information into three-dimensional space. Investigation of the energetic bases of these processes is a vast area for calorimetric studies. The importance and necessity of such studies has led to recent progress in calorimetric techniques, appearance of the super-sensitive reaction and the heat capacity microcalorimeters. Increase in sensitivity of calorimetric instruments was

*Paper based on a presentation at the 19th International Conference on Chemical Thermodynamics (ICCT-19), 30 July to 4 August 2006, Boulder, CO, USA. Other presentations are published in this issue, pp. 1345–1462.

‡Corresponding author: Tel.: +1 410-516-6532; E-mail: privalov@jhu.edu

required because most biological macromolecules are available in very limited amounts and because the intra- and intermolecular processes need to be studied in highly dilute solutions.

MICROCALORIMETRIC TECHNIQUES

The isothermal titration (reaction) microcalorimeters (ITCs) permit the measurement of heat effects from injecting small portions of one reagent into the dilute solution of another reagent (Fig. 1) and thus yield the enthalpy of their association, ΔH^a (for these techniques, see [1–3]). From serial injections of the titrant, one gets its binding isotherm and thereby also the association constant, K^a , which yields the Gibbs energy of association, $\Delta G^a = -RT\ln(K^a)$ and, by combining the two, the entropy factor of association, $T\Delta S^a = \Delta H^a - \Delta G^a$. Repeating this procedure at various temperatures gives the temperature dependence of the enthalpy, i.e., the heat capacity effect of association, $\Delta C_p^a = d(\Delta H^a)/dT$.

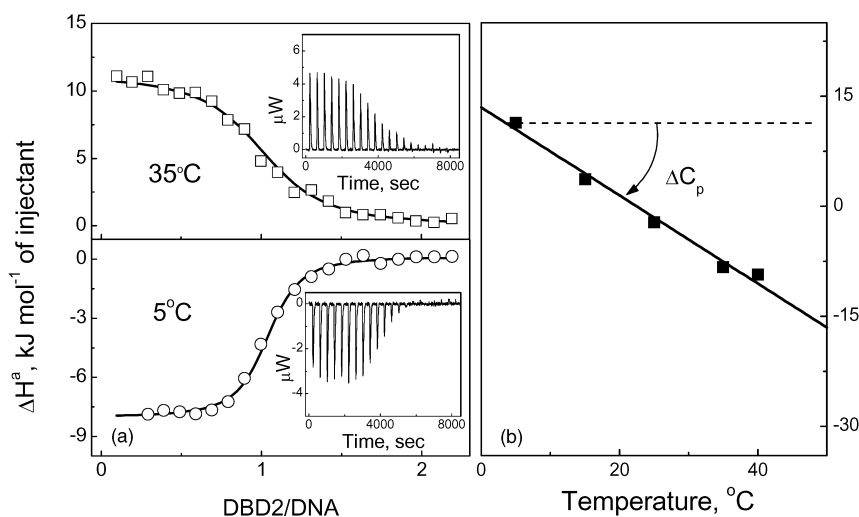


Fig. 1 (a) Calorimetric titration of a 16 bp DNA by the AT-hook peptide (DBD2) at two different temperatures using the ITC CSC 4200 with cell volume 1.25 ml, the volume of injected titrant being 10 μ l. (b) The dependence of the association enthalpy on temperature that determines the heat capacity effect of association. For details, see [4].

The differential scanning microcalorimeter (DSC), measures the small differences between the heat capacities of the standard liquid (solvent) and the studied solution over a broad temperature range (for this technique, see [5–7]). Using this apparent difference, $\Delta C_p^{\text{app}}(T)$, one can determine the partial specific heat capacity of the solute molecule, e.g., that of a protein:

$$C_p(T)^{\text{pr}} = C_p(T)^{\text{solv}} \times V(T)^{\text{pr}}/V(T)^{\text{solv}} - \Delta C_p^{\text{app}}(T)/m^{\text{pr}}(T) \quad (1)$$

where $C_p(T)^{\text{pr}}$ is the partial specific heat capacity of the protein at temperature T ; $V(T)^{\text{pr}}$ is the partial specific volume of the protein at T ; $V(T)^{\text{solv}}$ is the partial specific heat capacity of solvent and $m^{\text{pr}}(T)$ is the mass of protein in the calorimetric cell at this temperature (for details, see [5]). The partial heat capacity function of protein is one of the most important characteristics of its thermal properties (Fig. 2).

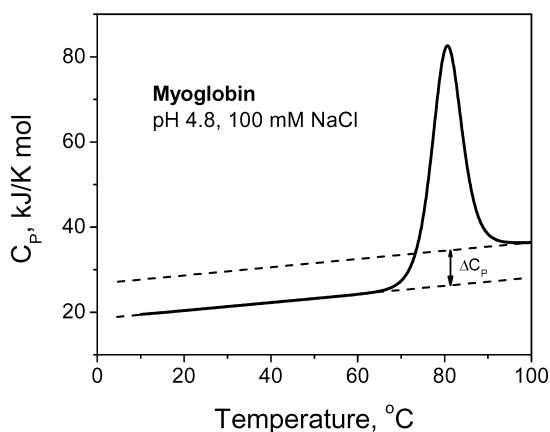


Fig. 2 The partial specific heat capacity of sperm whale metmyoglobin in 10 mM sodium acetate, pH 4.8, 100 mM NaCl. The net C_p/T function in molar terms: the dashed lines represent linear and parallel extrapolation of the heat capacities of the native and unfolded states [10].

Thermal properties of proteins

One can see from the example of myoglobin (Fig. 2) that globular proteins do not change much upon heating up to some critical temperature, but then unfold with extensive heat absorption resulting in a significant increase of their heat capacity, $\Delta C_p = C_p^U - C_p^F$. The area of the heat absorption peak gives the enthalpy of unfolding, ΔH . In the case of small globular proteins (molecular mass < 20 kDa) this enthalpy was found to be in correspondence with the van't Hoff enthalpy determined from the sharpness of heat absorption profile, which means that the temperature-induced unfolding of these proteins is a highly cooperative process [8,9]. It looks, therefore, as if the macromolecule jumps reversibly from the folded state to the unfolded state without noticeable intermediates. In that case, the equilibrium constant at the mid-transition temperature, T_t , equals 1, the Gibbs energy difference between the states at that temperature is zero, $\Delta G_t = -RT \ln(K_t) = \Delta H_t - T_t \Delta S_t = 0$, and thus the unfolding entropy at that temperature is $\Delta S_t = \Delta H_t / T_t$.

Since according to Kirchhoff's relation, $\partial (\Delta H) / \partial T = \Delta C_p$, and $\partial (\Delta S) / \partial T = \Delta C_p / T$, and assuming that the difference between the heat capacities of the folded and unfolded forms of the protein does not depend on temperature, i.e., $\Delta C_p \approx \text{constant}$, one finds:

$$\Delta H(T) = \Delta H_t + \int_{T_t}^T \Delta C_p dT \approx \Delta H_t + \Delta C_p (T - T_t) \quad (2)$$

$$\Delta S(T) = \frac{\Delta H(T_t)}{T_t} + \int_{T_t}^T \frac{\Delta C_p(T)}{T} dT \approx \Delta H_t / T_t + \Delta C_p \cdot \ln(T / T_t) \quad (3)$$

Their combination gives for the Gibbs energy:

$$\begin{aligned} \Delta G(T) &= \Delta H(T) - T \Delta S(T) \approx \Delta H_t [T_t - T] / T_t - \Delta C_p [(T_t - T) + \ln(T / T_t)] \\ &\approx \Delta H_t [T_t - T] / T_t - \Delta C_p [(T_t - T) / 4T]^2 \end{aligned} \quad (4)$$

According to these equations, the enthalpy and entropy of protein unfolding are continuously increasing functions of temperature, while the Gibbs energy, which is the small difference between the enthalpy and the entropy factor, is a function with a maximum which crosses zero at two different temperatures (Fig. 3). Since the Gibbs energy represents the work required to transfer protein from the

folded state to the unfolded state, it is regarded as a measure of protein stability. It follows, therefore, that a protein should unfold not only upon heating but also upon cooling. According to this prediction, protein unfolding upon cooling should typically take place at temperatures below the freezing point of aqueous solutions—a serious obstacle to its observation. However, by super-cooling aqueous solutions, its occurrence was demonstrated first on myoglobin and apomyoglobin [10,11] (Fig. 4) and then shown that “cold denaturation” is specific for all globular proteins (for review, see [12]). The most intriguing property of cold denaturation is that it proceeds with heat release.

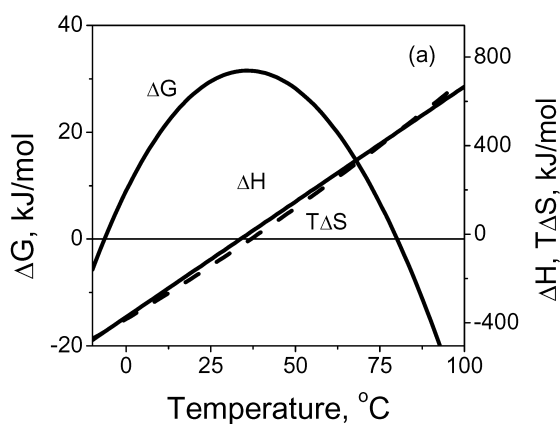


Fig. 3 The temperature dependence of the unfolding enthalpy, the entropy factor, $T\Delta S$, and their difference, i.e., the Gibbs energy, calculated assuming that the heat capacity increment of unfolding does not depend on temperature. For details, see [9–11].

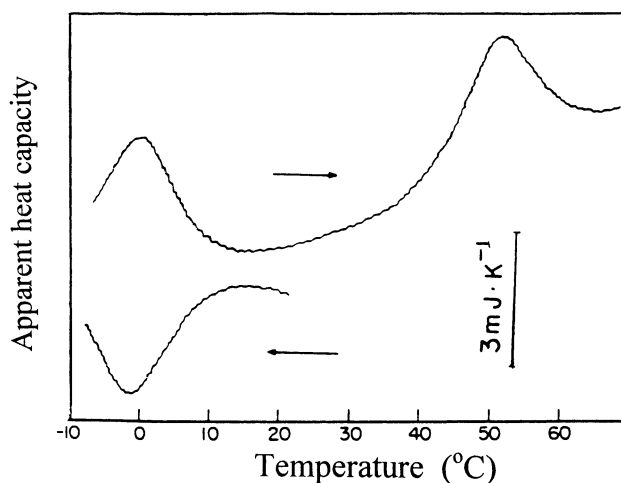


Fig. 4 The heat effects recorded by DSC upon cooling an apomyoglobin solution (pH 4.8) from room temperature to $-10\text{ }^{\circ}\text{C}$ and its consequent heating to $80\text{ }^{\circ}\text{C}$ [10,11].

It thus followed that the thermodynamic formalism correctly predicted the thermal properties of proteins. Not all was clear, however, concerning these properties. Firstly, it was unclear why protein unfolding upon cooling proceeds with heat release and entropy decrease (i.e., with an *increase of order*) in contrast to the unfolding upon heating that proceeds with heat absorption and an entropy increase (i.e., an *increase in disorder*). Secondly, it was unclear how far the enthalpy and entropy of protein un-

folding could increase with temperature rise. They cannot increase indefinitely, as follows from eqs. 2 and 3, since the number of interactions disrupted upon unfolding are limited in proteins and the extent of disorder reached by the polypeptide chains of proteins upon unfolding is also limited.

In fact, an unlimited increase of enthalpy and entropy with temperature rise is only the consequence of assuming that the heat capacity difference between the folded and unfolded states of a protein does not depend on temperature. Detailed studies of the temperature-induced unfolding of small compact globular proteins showed that it does in fact depend on temperature (Fig. 5): while the heat capacity of the native state is a linearly increasing function of temperature, the heat capacity of the unfolded state is a curved function which crosses the heat capacity of the native state at high temperatures [13,14]. Thus, the difference between the heat capacities of the native and the unfolded states of proteins decreases with temperature rise and vanishes at about 120 °C. If so, the enthalpy of protein unfolding/refolding should be an asymptotic function of temperature, reaching some constant level at this temperature (Fig. 6, curves 1 and 2).

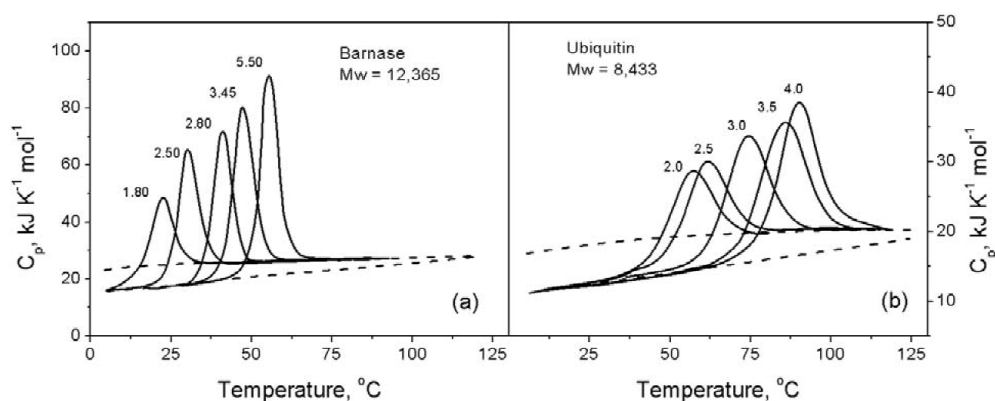


Fig. 5 The partial heat capacity functions of barnase and ubiquitin at different pHs showing that the heat capacity difference between the unfolded and folded states decreases with temperature rise and disappears at about 120 °C. For details, see [13,14].

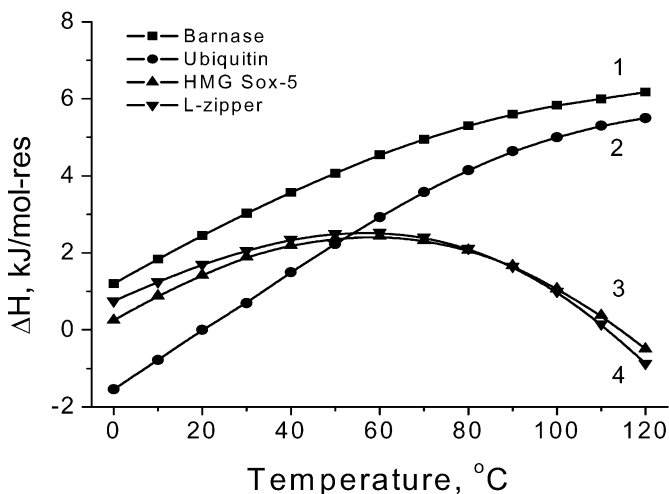


Fig. 6 The temperature dependencies of the enthalpy of unfolding of barnase (curve 1), ubiquitin (curve 2), DBD HMG Sox-5 (curve 3), and leucine zipper, L-zip (curve 4) obtained on linear extrapolation of the initial heat capacity function, presumably describing the native folded state of protein.

Marginally stable proteins

With increase of the number of calorimetrically studied proteins, it appeared that, while the heat capacity of the unfolded state of all proteins is rather similar, the heat capacity of the native states varies considerably: in some cases, the initial slope is rather steep and its extrapolation crosses the heat capacity of the unfolded state not at 120 °C but at much lower temperatures (Fig. 7) [15,16]. Thus, the difference between the heat capacities of the folded and unfolded states of such a protein changes sign at the “crossing-over” temperature. According to eq. 2, up to this temperature the enthalpy of unfolding should be an increasing function of temperature and above this temperature it should decrease (Fig. 6, curves 3 and 4).

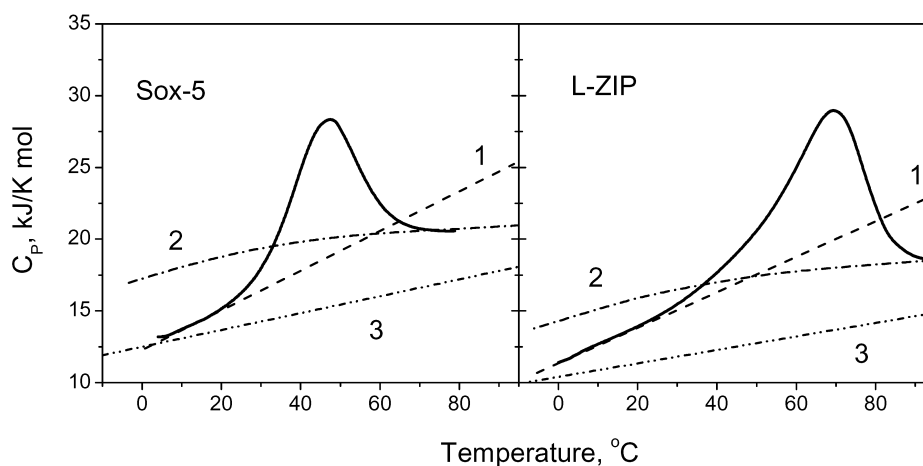


Fig. 7 The partial heat capacity of the DBD HMG Sox-5 and the leucine zipper L-ZIP GCN4. The heat capacity of the initial state in those cases increases very steeply, and its linear extrapolation crosses the heat capacity of the unfolded state at temperatures much below 120 °C. In both cases, line 1 represents linear extrapolation of the initial heat capacity function of protein; line 2 represents the calculated heat capacity function of the unfolded protein; line 3 shows the standard heat capacity function of the fully folded protein estimated using the partial specific heat capacity of BPTI. For details, see [15,16].

Further studies showed that such a steep initial increase of the heat capacity function is not an exception in proteins and its slope varies over a rather broad range [17]. The lowest slope has BPTI, a compact protein with an extremely rigid structure stabilized by three S-S cross-links (Fig. 8). For proteins with higher B factors (i.e., with more loose structures), the initial slope is greater. It thus looks as if the initial slope of the heat capacity function reflects temperature-induced intensification of fluctuations of protein structure (i.e., indicates that heating results in the accumulation of energy in flickering local unfoldings). If so, this excess heat effect should be taken into account when estimating the enthalpy of protein unfolding. The question is then how to determine this excess heat. This needs the heat capacity function of a fully folded protein to be taken as a baseline in estimating excess heat effects.

As a standard heat capacity of fully folded protein, one can perhaps take the partial specific heat capacity function of BPTI (Fig. 7, dashed-and-dotted lines). Using this standard as a baseline, one can determine the enthalpy and entropy functions of protein unfolding by integrating the excess heat effect [17]:

$$\Delta_N^U H(T) = \int_{T_o}^{T_{\max}} [C_P^{\text{pr}} - C_P^{\text{F}}] dT - \int_{T_{\max}}^T [C_P^{\text{U}} - C_P^{\text{F}}] dT \quad (5)$$

$$\Delta_N^U S(T) = \int_{T_o}^{T_{\max}} [C_P^{\text{pr}} - C_P^{\text{F}}] d \ln T - \int_{T_{\max}}^T [C_P^{\text{U}} - C_P^{\text{F}}] d \ln T \quad (6)$$

Here, C_P^{pr} is the partial heat capacity of the protein, C_P^{F} is the partial heat capacity of the completely folded state of the protein calculated using the specific heat capacity of BPTI, C_P^{U} is the partial heat capacity of the unfolded polypeptide chain calculated by summing up the heat capacities of the amino acid residues constituting the protein, and T_{\max} is the temperature to which the protein was heated in the calorimetric experiment and at which it is assumed to be completely unfolded. Such analysis of the heat capacity functions of calorimetrically studied globular proteins showed that in all cases their specific enthalpy and entropy of unfolding (calculated per gram or per mole of residues) approach asymptotically to almost the same limits at 120 °C (Fig. 9).

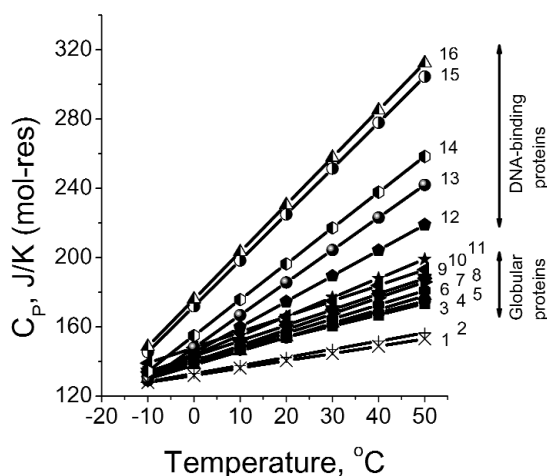


Fig. 8 The initial slopes of the partial specific heat capacity functions of a number of proteins: 1, anhydrous chymotrypsinogen; 2, BPTI; 3, barnase; 4, myoglobin; 5, lysozyme; 6, cytochrome c; 7, ubiquitin; 8, ubiquitin; 9, RNase T1; 10, RNase A; 11, engrailed homeodomain; 12, Mat $\alpha 2$ homeodomain; 13, Antennapedia homeodomain; 14, DBD HMGD-74; 15, L-zipper GCN4; 16, DBD HMG Sox5; 17, Zn-finger TFIIIA; 18, DBD HMG NHP6A; 19, DBD HMG Sry; 20, DBD HMG Lef-1. For refs., see [17].

The question then is: Upon temperature decrease, why do the enthalpy and entropy reduce from the values reached at 120 °C and even change sign at low temperatures? The internal interactions and conformational entropy should not decrease so much on lowering the temperature, and, moreover, they should not change sign at low temperatures. These effects can be caused only by the surrounding medium (i.e., the water).

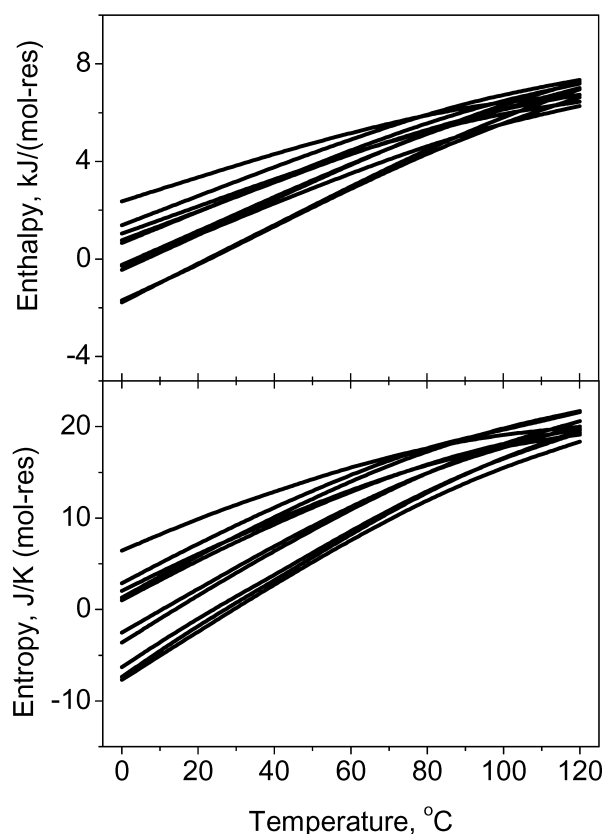


Fig. 9 The enthalpy and entropy functions of unfolding of the number of proteins listed in Fig. 7. These functions asymptotically approach common limits at 120 °C [17]. At this temperature, the average unfolding enthalpy of all globular proteins appears to be (6 ± 1) kJ/(mol-res), and their averaged unfolding entropy is about (18 ± 5) kJ/K (mol-res).

Hydration effects

The unique native structure of proteins is actually determined by the energy balance of various types of interaction between individual protein groups and between these groups and water. Interaction of the various groups of proteins with water is thus of crucial importance (for reviews, see [18,19]). The results from many laboratories are summarized in Figs. 10a and 10b. Transfer of polar groups into water proceeds with negative enthalpy and entropy (i.e., positive $-T\Delta S$), and this is understandable since polar groups form hydrogen bonds with the polar water molecules and dampen the motion of the surrounding water. The enthalpy change overbalances the entropy factor, so that the Gibbs energy of transfer is negative at all temperatures (i.e., the process is thermodynamically favorable). In consequence, polar groups are hydrophilic.

It was most surprising, however, to find that polar water molecules interact also with apolar groups: the hydration of apolar groups likewise results in a decrease of the enthalpy and entropy (see Fig. 10b). This is explained by an increase in the interactions between water molecules and by water ordering around the apolar groups. In this case, the entropy factor overbalances the enthalpy and the resulting positive Gibbs energy manifests itself in the low solubility of apolar groups. The tendency of apolar groups to aggregate is usually described as a “hydrophobic force” repulsing them from water [20], a force usually assumed to be entirely entropic. This is true, however, only at room temperature (~ 20 °C), where the positive dehydration enthalpy of apolar groups compensates exactly the negative

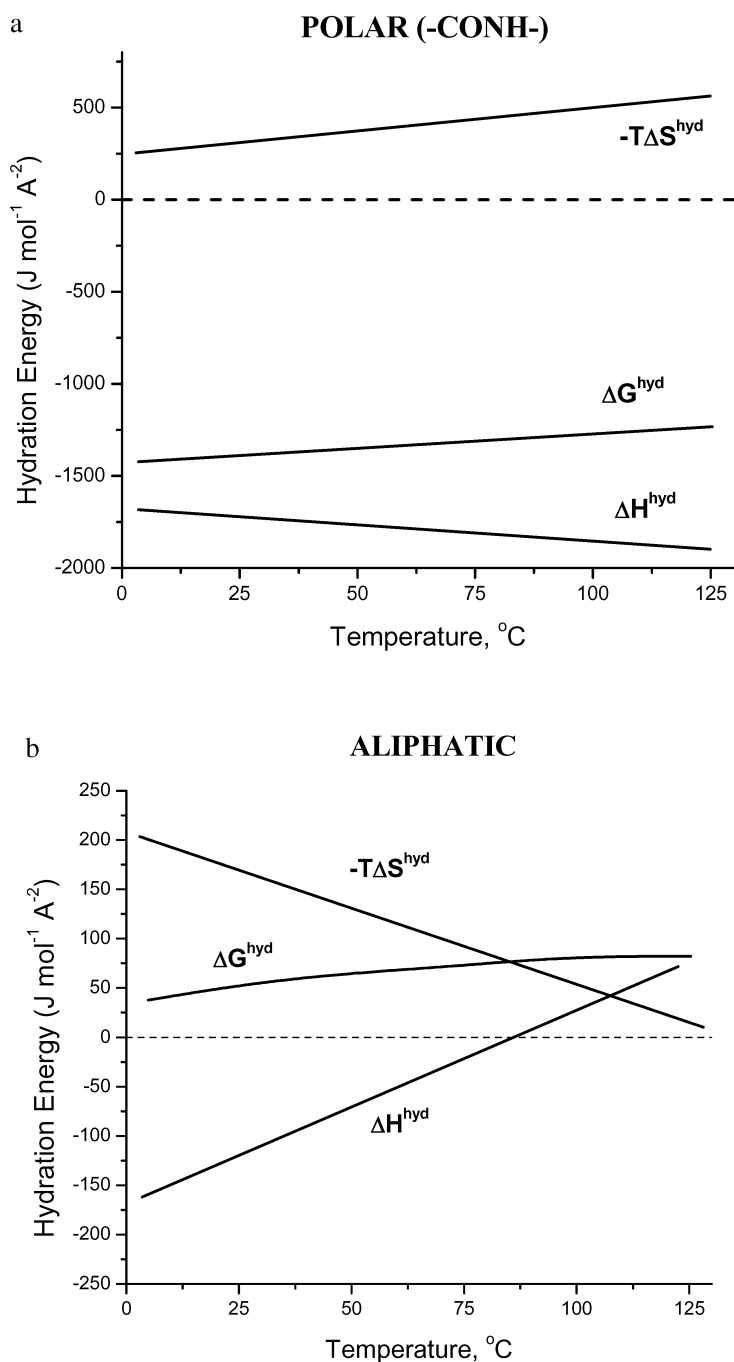


Fig. 10 The hydration effects of polar (a) and apolar (b) groups in proteins. As a polar group is considered, the peptide unit which is most populated group in proteins. For details, see [18].

enthalpy of forming van der Waals interactions between these groups [18,19]. This hydrophobic force is assumed to be responsible for the formation of the apolar core of globular protein [20].

It is notable that with temperature increase the negative enthalpy and entropy of hydration of polar groups increase in magnitude, in contrast to the hydration of apolar groups for which they decrease in magnitude (Figs. 10a and 10b). It follows, therefore, that the heat capacity effect (ΔC_p) of hydrating polar groups is negative, while for apolar groups ΔC_p is positive. Correspondingly, the disruption of hydrophobic contacts results in a heat capacity increase. This is just what we see upon unfolding globular proteins, which results in exposure of the internal apolar groups that formed their hydrophobic core [18].

The overall hydration effects upon protein unfolding can be calculated if we know the change of water-accessible surface area, ΔASA_i , of every type of protein group, i , upon unfolding. Then, using the surface-normalized intrinsic hydration effects of this group, $\Delta \hat{H}_i^{\text{hyd}}$ and $\Delta \hat{S}_i^{\text{hyd}}$, determined using model compounds, one can estimate the hydration effects upon protein unfolding [18]:

$$\Delta H^{\text{hyd}} = \sum \Delta \hat{H}_i^{\text{hyd}} \times \Delta ASA_i \text{ and } \Delta S^{\text{hyd}} = \sum \Delta \hat{S}_i^{\text{hyd}} \times \Delta ASA_i \quad (7)$$

The hydration entropies of polar and apolar groups upon unfolding of various globular proteins are shown in Fig. 11. We see that the hydration entropies of the apolar groups vanish at 120 °C but the hydration entropies of the polar groups become even larger at that temperature, varying for different proteins between -25 and -37 J/K (mol-res).

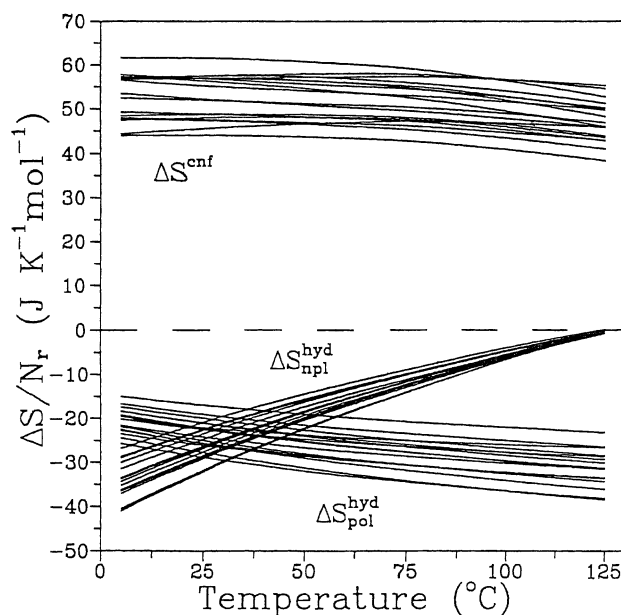


Fig. 11 Temperature dependence of the entropies of hydration for polar, $\Delta S_{\text{pol}}^{\text{hyd}}$, and apolar groups, $\Delta S_{\text{apol}}^{\text{hyd}}$ and of the conformational entropies of unfolding, ΔS^{conf} of various globular proteins [18].

Conformational entropy of protein unfolding

The calorimetrically determined entropies of protein unfolding at 120 °C, (20 ± 3) J/K (mol-res) (Fig. 9), include the hydration entropies of polar groups, $-(31 \pm 7)$ J/K (mol-res) (Fig. 11). If we exclude this hydration effect from the calorimetrically determined entropies of unfolding at this temperature, we find that the conformational entropy of protein unfolding should be about (51 ± 10) J/K (mol-res).

The change in conformational entropy of protein unfolding was calculated by many scientists analyzing theoretically the possible increase of conformational freedom of a polypeptide chain in a vac-

uum and led to the conclusion (Table 1) that it should be around (40 ± 10) J/K (mol-res). Notwithstanding the somewhat simplified assumptions used in these calculations and the rather large extrapolation of the hydration effects up to 120 °C, the conformational entropy values obtained by these two approaches are rather close. This shows that we more or less understand the forces responsible for the formation of unique native structures of proteins. Nevertheless, there are still some unsolved problems:

- (a) We do not understand the mechanism of cooperation between the forces involved in protein folding which makes the unfolding/refolding process so sharp. It is precisely because of this cooperation that proteins have definite and stable enough structures at temperatures close to physiological. Indeed, the Gibbs energy of stabilization of the native structure of a protein consisting of 100 amino acid residues is about 30 kJ/mol at physiological temperatures, i.e., each residue is contributing only 0.3 kJ/mol, which is much below the thermal energy at that temperature (2.4 kJ/mol).
- (b) Not everything is clear with regard to the hydration effects. In particular, why do apolar groups order the water? Why is the heat capacity effect of hydrating apolar group positive, in contrast to the negative heat capacity effect of hydrating polar group? There are many speculations on this matter, but no quantitative theory describes these effects. This, however, is not entirely surprising since we still do not have a good theory for the pure water.

Table 1 The theoretically calculated entropy increase upon unfolding of globular proteins in vacuum.

References	Main chain	Side chain	Total
Nicholls, Sharp, Honig [21]		18	
Sterenberg, Chickos [22]		14	
Doig, Sterenberg [23]		14	
Wang et al. [24]	21	18	41
D'Aquino et al. [25]	18	14	32
Honig, Yang [26]	29	18	52

Temperature-induced changes in proteins

As follows from the above, the temperature-induced change of small globular proteins is not a simple process in all cases: before the gross conformational transition of a protein to the unfolded state takes place, some changes in its structure occur that are associated with an increase of heat capacity. In some cases, these changes are gradual and might be explained by an intensification of local fluctuations of protein structure. Upon reaching a critical temperature, these fluctuations might start to correlate and their final synchronization result in the cooperative unfolding of whole protein [27]. The segment of protein structure over which such cooperation takes place is usually called the “cooperative domain”. Large proteins consist of several cooperative domains and, correspondingly, they unfold in several discreet steps [28]. However, small proteins ($M_w < 20$ kDa) are usually regarded as consisting of a single domain [9].

The question then is: How independent are the fluctuations of a single-domain protein below this critical temperature? A very steep increase of the heat capacity, as observed in the case of some of the proteins, indicates that fluctuations of their structure might be synchronized within certain clusters of

residues. These clusters might be regarded as cooperative subdomains of the protein structure and be revealed by a deconvolution analysis of the heat capacity profile of the protein (i.e., its approximation by a number of two-state transitions). An unequivocal solution of this task is possible since the enthalpy and temperature are conjugate-extensive and -intensive variables, and their functional dependence includes all the information on the macroscopic states which are realized over the considered temperature range [29].

The results of such an analysis of the heat capacity functions of some marginally stable globular proteins are shown in Fig. 12. All of these proteins are rather small: less than 12 kDa molecular mass and usually regarded as single domain. We see, however, that the structure of these proteins does not represent a single monolith, but includes two/three cooperative subdomains, some of which are partly or even completely unfolded at room temperature.

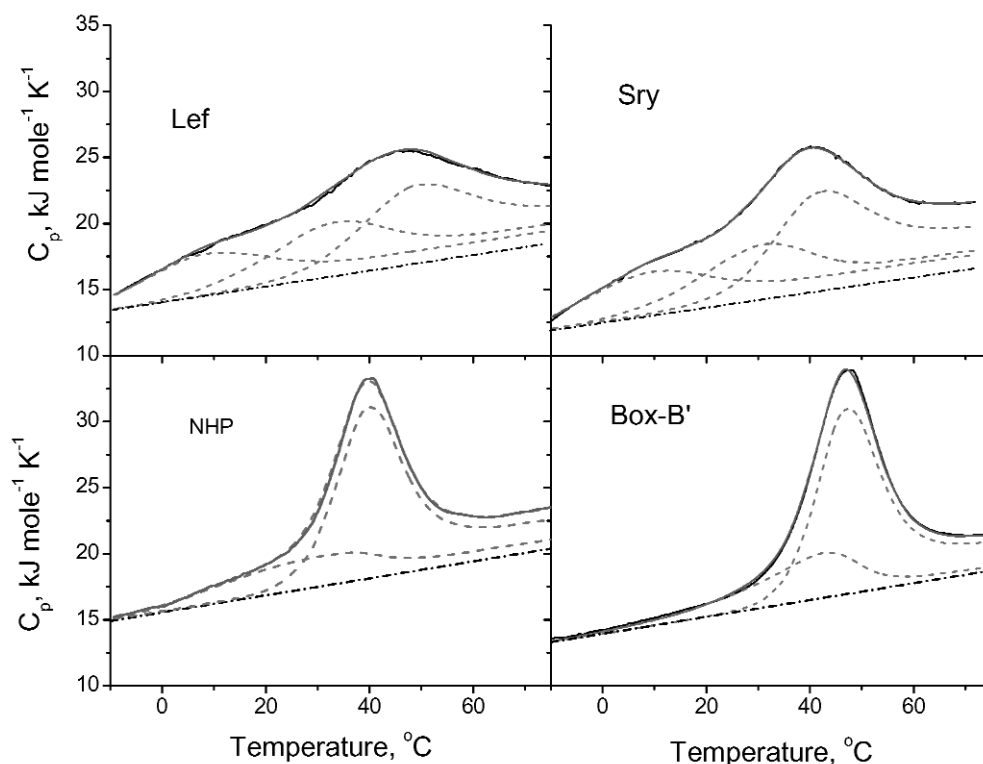


Fig. 12 Deconvolution analysis of the heat capacity functions of the DBDs of HMG Lef-1, HMG Sry HMG NHP6A, and HMG Box-B' showing that these small proteins (M_w about 10 kDa) unfold in several cooperative steps (i.e., their structure is subdivided into two/three cooperative subdomains). For details, see [30].

Formation of protein complexes

Inspection of the initial heat capacity functions presented in Fig. 8 shows that there are two groups of proteins which differ in their heat capacity slopes: low slopes are characteristic of proteins which work independently, whereas steep slopes are specific for proteins which work in complex with other macromolecules, in particular with DNA. It appears that DNA-binding proteins (domains) have more flexible structures, and this, perhaps, is necessary to allow them to more efficiently enwrap their sequence-specific partner [30]. Let us consider some of these proteins and their interaction with target DNAs.

One of the largest families of DNA-binding proteins involved in regulating transcription is the high mobility group of proteins, containing the so-called HMG box, i.e., the DNA-binding domain

(DBD) [30]. The heat capacity functions of the DBDs of four representative of this family are shown in Fig. 12. Upon heating these DBDs, the excess heat absorption starts from very low temperatures. It thus appears that at room temperature, and the more so at physiological temperatures (37 °C), these proteins are partly unfolded. However, upon association with DNA in an ITC experiment, they refold to form complexes which melt cooperatively at substantially higher temperatures than does the unbound protein (Fig. 13). Refolding of the protein component results in release of heat, and the magnitude of this refolding depends on the temperature at which it takes place (i.e., the temperature at which the ITC experiment is conducted).

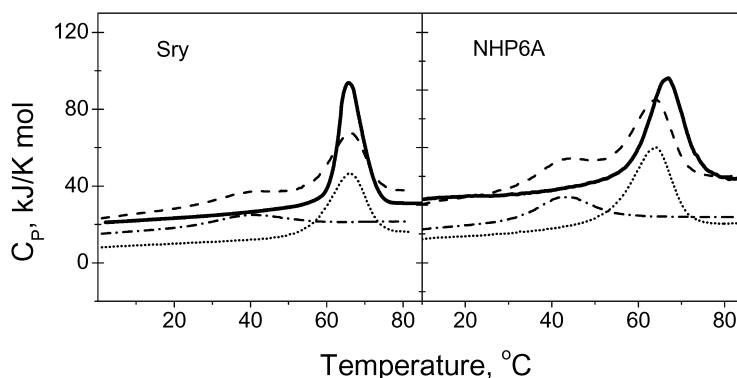


Fig. 13 The partial molar heat capacities functions of the free DBDs of Sry and NHP6a, their free target DNA duplexes, and their 1:1 complexes. The dashed lines show the sum of the heat capacities of the free protein and the free DNA. For details, see [30].

Figure 14a shows the apparent enthalpies of binding the SRY and NHP6A DBDs to their cognate DNA duplexes, measured by ITC at different temperatures. The observed negative enthalpies change nonlinearly with temperature, becoming even more negative because the heat evolved on refolding increases nonlinearly with temperature. Using the DSC-measured heat capacity functions of all the components of binding reaction (in this case, free protein, free DNA, and their complex) one can determine the heat of protein refolding upon binding to DNA at any desired temperature (for details, see [31]). Correcting the apparent binding enthalpies for the heat of refolding, we get the enthalpy of association of the fully folded protein with DNA. This correction significantly changes the magnitudes of the enthalpy and linearizes its dependence on temperature, which defines the heat capacity effect of binding (Fig. 14b). Correction of the calorimetrically measured thermodynamic characteristics for protein refolding is especially important in analyzing the energetic bases of the complexes. This is because in the structure of the complex determined by crystallography or NMR, the protein is in the folded state (i.e., the interface is formed by fully folded protein). Thus, the structural parameters of the complex (e.g., the change in water-accessible surface area of the apolar and polar groups at the interface) will correlate only with binding characteristics corrected for protein refolding.

It is worth noting that the Gibbs energy of binding does not require correction for refolding since the free energy of temperature-induced conformational changes of proteins at temperatures around ambient are usually small and may be neglected. Correspondingly, the binding entropy factor (derived from the difference between the Gibbs free energy and the corrected binding enthalpy) is also corrected for protein refolding.

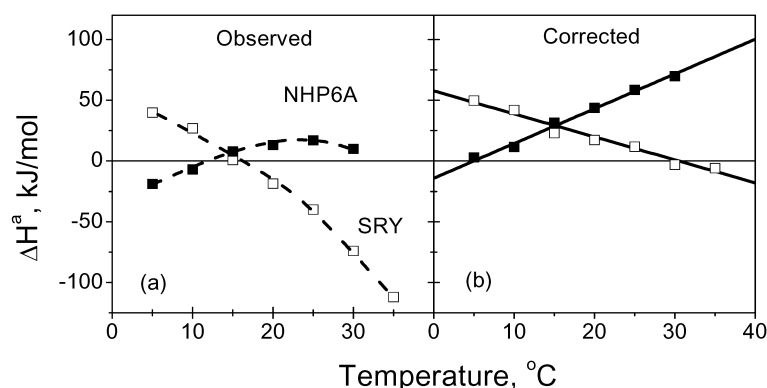


Fig. 14 (a) The ITC-measured association enthalpies of the DBDs of Sry and NHP6A HMG proteins with their optimal DNAs plotted as functions of temperature. (b) The association enthalpies corrected for the refolding of protein upon binding. For details, see [30].

Energetic signatures of protein binding to DNA

The DNA-binding proteins interact specifically with either the major or minor grooves of DNA, and it is remarkable that binding to the minor groove usually takes place at AT-rich sequences. One can then ask if there are qualitative differences in the forces driving protein binding to the different grooves of DNA. Figure 15 presents Gibbs energies, enthalpies, and entropies of binding of various proteins (corrected for refolding) to their target DNAs at 20 °C in the presence of 100 mM NaCl (for details, see [30–33]).

It appears that there are no dramatic differences in the Gibbs free energies of protein binding to the major and minor grooves of DNA (Fig. 15a). Particularly striking, therefore, is the systematic qualitative difference in binding enthalpies for the two grooves (Fig. 15b): the enthalpies of binding to the minor groove are always positive, while the enthalpies of binding to the major groove are invariably negative. Since the Gibbs free energies of binding at 20 °C do not vary greatly, the enthalpy differences are essentially balanced by the entropic factors. Indeed, the positive entropy factor (i.e., $T\Delta S$) of minor groove binding is significantly larger in magnitude than that of major groove binding (Fig. 15c). Thus, it appears that binding of protein to the minor groove is entropy-driven, in contrast to major groove binding, which is enthalpy-driven.

Although binding of protein to the narrow minor groove of AT-rich DNA in many cases proceeds with significant deformation of the DNA, this does not explain the observed drastic difference in the enthalpy of binding to the minor and major grooves because not all the minor groove binders presented in Fig. 14 bend DNA [33]. Moreover, one might expect that the work for DNA bending would manifest itself in the Gibbs energy of binding, but this is not the case. The large positive enthalpy of protein binding to DNA most probably results from dehydration of the protein/DNA interface [33]. It appears that, unlike the major groove, the enthalpy of dehydrating the minor groove far exceeds the favorable enthalpic contributions from newly formed binding interactions.

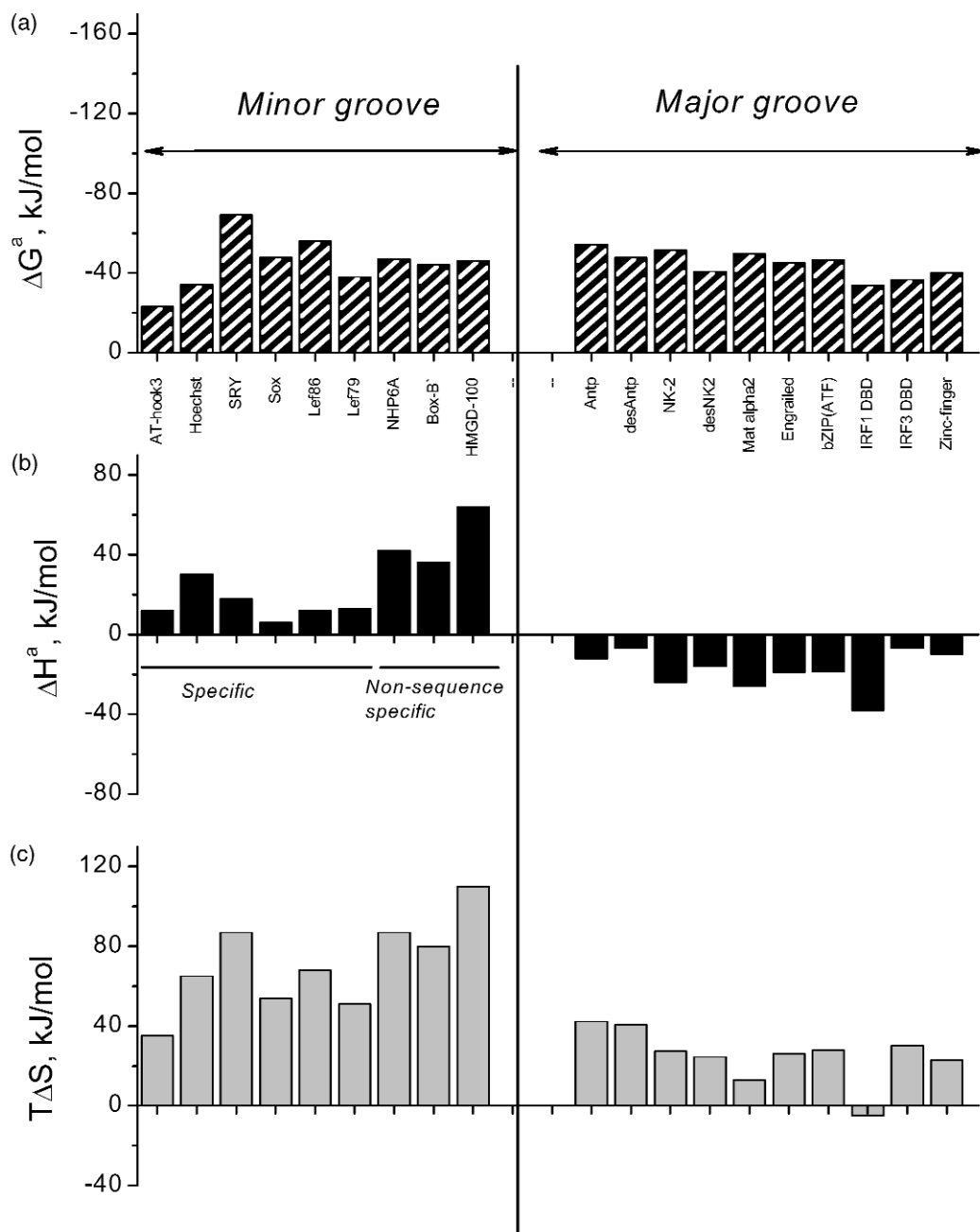


Fig. 15 The thermodynamic parameters of association for various DBDs to the major and minor grooves of DNA at 20 °C in the presence of 100 mM NaCl: (a) Gibbs energies of association; (b) enthalpies of association; (c) entropy factors of association. For details, see [33].

Hydration of the DNA grooves

The presence of a spine of well-ordered water molecules in the minor groove of AT-rich DNA sequences has been shown by high-resolution X-ray crystallography and neutron diffraction [34,35]. The minor groove is unusually narrow in AT stretches, and a primary shell of water molecules runs across

the groove, bridging acceptor sites on adjacent AT/TA base pairs (Fig. 16). In the major groove, the distribution of H-bond donors and acceptors exhibits greater irregularity than in the minor groove and no regular water superstructure has been identified.

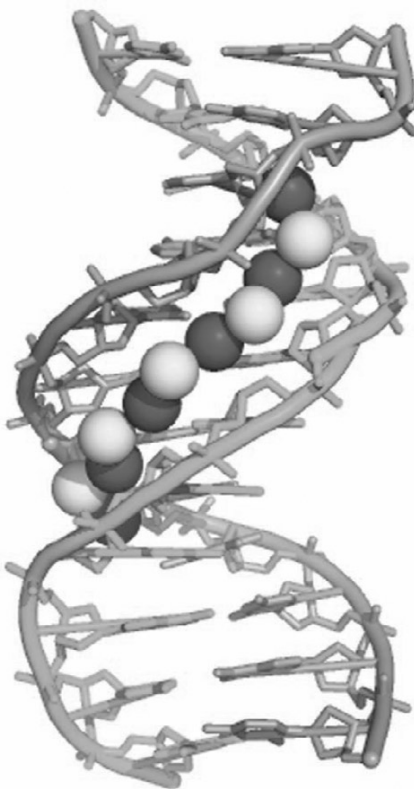


Fig. 16 Display of the spine of water in the minor groove of the X-link dodecamer CGCAATTCGCG generated from the coordinates of NDB accession number bd0008.

If the water hydrating the minor groove of AT-rich DNA is more ordered than the water hydrating the major groove, its removal is expected to require a greater enthalpy and thereby provides a larger entropy increase than removal of water hydrating the major groove. However, this is not a manifestation of a hydrophobic force upon protein binding to DNA since water ordering in the AT-rich minor groove is determined not by the apolar groups of the DNA but by the regular arrangement of its polar groups that stabilize an ice-like organization of the water in this groove.

CONCLUSION

In studying the energetics of protein unfolding/refolding, attention has been usually concentrated on the most pronounced stage of this process, the cooperative transition. This is because this stage takes place over a relatively short range of external variables (e.g., temperature), it can be observed by a variety of experimental techniques and can easily be analyzed and specified quantitatively in thermodynamic parameters. The physics of this cooperative process of protein unfolding/refolding does indeed represent an interesting and important problem. This interest, however, has almost completely obscured other, more subtle changes of protein structure under the influence of changing environmental conditions, particularly temperature. These changes are less impressive than those of the gross cooperative

conformational transition because they are not so pronounced; correspondingly, their experimental investigation is more difficult and the theoretical analysis is more complicated. However, they take place at physiological temperatures and may, therefore, have functional significance.

The observed predenaturational changes show that protein structure is not uniform in stability and flexibility: it can be rather unstable as a whole or might have unstable parts which intensively fluctuate under the considered conditions. The stability and flexibility of the protein or its subdomains can change upon association with its functional partners: another protein, nucleic acid, or specific ligand. These dynamic properties of protein structure are now of particular interest because they determine the ability of a protein to bind its partner, which is the most basic function of a protein. Without studying the thermodynamics of these subtle changes of protein structure, one cannot get a proper understanding of the energetic basis of their complexes, and consequently of the mechanism of their formation and functioning.

The other important aspect of thermodynamics of biological macromolecules concerns the role of water in the formation of their spatial structure. It appears that water plays an important role not only in the thermodynamics of protein folding, but also in the formation of protein complexes, particularly with DNA. Understanding the state of water in the vicinity of the groups of proteins and in the DNA grooves is, therefore, of paramount importance for understanding the mechanisms of folding proteins and their association with DNA. As stated above, hydration effects in proteins have been extensively studied, but there are still several unsolved questions. As for the hydration of DNA, this has not yet been studied in detail and it is already clear that it is very different for the minor and major grooves, particularly at AT-rich sequences. With raised interest in the mechanisms of gene regulation, the state of water in the DNA grooves and its role in the process of protein-DNA recognition deserves serious attention.

ACKNOWLEDGMENTS

I thank all coauthors of the results discussed in this lecture: Ken Breslauer, Mair Churchill, Colyn Crane-Robinson, Anatoly Dragan, Yuri Griko, Patrick Wintrobe, Elena Makeyeva, George Makhataдзе, Conceição Minetti, Christopher Read, and David Remeta. I also acknowledge the financial support of NIH GM48036 and NSF MCB 0519381.

REFERENCES

1. I. R. McKinnon, L. Fall, A. Parody-Morreale, S. J. Gill. *Anal. Biochem.* **139**, 134 (1984).
2. K. J. Bresslauer, E. Freire, M. Straume. *Methods Enzymol.* **211**, 533 (1992).
3. I. Wadso. *Thermochim. Acta* **250**, 285 (1995).
4. A. I. Dragan, J. R. Liggins, C. Crane-Robinson, P. L. Privalov. *J. Mol. Biol.* **327**, 392 (2003).
5. P. L. Privalov, S. A. Potekhin. *Methods Enzymol.* **131**, 4 (1986).
6. P. L. Privalov, V. V. Plotnikov. *Thermochim. Acta* **139**, 257 (1989).
7. G. P. Privalov, V. Kavina, E. Freire, P. L. Privalov. *Anal. Biochem.* **232**, 79 (1995).
8. P. L. Privalov, N. N. Khechinashvili. *J. Mol. Biol.* **86**, 665 (1974).
9. P. L. Privalov. *Adv. Protein Chem.* **33**, 167 (1979).
10. P. L. Privalov, Y. V. Griko, S. Y. Venyaminov, V. P. Kutysenko. *J. Mol. Biol.* **190**, 487 (1986).
11. Y. V. Griko, P. L. Privalov, S. Y. Venyaminov, V. P. Kutysenko. *J. Mol. Biol.* **202**, 127 (1988).
12. P. L. Privalov. *Crit. Rev. Biochem. Mol. Biol.* **25**, 281 (1990).
13. Y. V. Griko, G. I. Makhataдзе, P. L. Privalov, R. W. Hartley. *Protein Sci.* **3**, 669 (1994).
14. P. L. Wintrobe, G. I. Makhataдзе, P. L. Privalov. *Proteins: Struct., Funct., Genet.* **18**, 246 (1994).
15. C. Crane-Robinson, C. M. Read, P. D. Cary, P. C. Driscoll, A. I. Dragan, P. L. Privalov. *J. Mol. Biol.* **281**, 705 (1998).
16. A. I. Dragan, P. L. Privalov. *J. Mol. Biol.* **321**, 891 (2002).
17. P. L. Privalov, A. I. Dragan. *Biophys. Chem.* **126**, 16 (2007).

18. G. I. Makhatadze, P. L. Privalov. *Adv. Protein Chem.* **47**, 307 (1995).
19. P. L. Privalov, S. J. Gill. *Adv. Protein Chem.* **39**, 193 (1988).
20. W. Kauzman. *Adv. Protein Chem.* **14**, 1 (1959).
21. A. Nicholls, K. A. Sharp, B. Honig. *Proteins: Struct., Funct., Genet.* **11**, 281 (1991).
22. M. J. E. Sterenberg, J. S. Chickos. *Protein Eng.* **7**, 149 (1994).
23. A. J. Doig, J. M. Sterenberg. *Protein Sci.* **4**, 2247 (1995).
24. J. Wang, Z. Szewczuk, S. Yue, Y. Tsuda, Y. Konishi, E. Purisima. *J. Mol. Biol.* **253**, 473 (1995).
25. J. A. D'Aquino, J. Gomez, H. J. Hisler, K. H. Lee, L. M. Amzel, E. Freire. *Proteins: Struct., Funct., Genet.* **25**, 143 (1996).
26. B. Honig, A. S. Yang. *Adv. Protein Chem.* **46**, 25 (1995).
27. A. Ikegami. *Biophys. Chem.* **6**, 117 (1977).
28. P. L. Privalov. *Adv. Protein Chem.* **35**, 1 (1982).
29. E. Freire, R. L. Biltonen. *Biopolymers* **13**, 463 (1978).
30. A. I. Dragan, C. M. Read, E. N. Makeyeva, E. I. Milgotina, C. Crane-Robinson, P. L. Privalov. *J. Mol. Biol.* **343**, 371 (2004).
31. P. L. Privalov, I. Jelesarov, C. M. Read, A. I. Dragan, C. Crane-Robinson. *J. Mol. Biol.* **294**, 997 (1999).
32. A. I. Dragan, Z. Li, E. N. Makeyeva, E. I. Milgotina, Y. Liu, C. Crane-Robinson, P. L. Privalov. *Biochemistry* **45**, 141 (2006).
33. P. L. Privalov, A. I. Dragan, C. Crane-Robinson, K. J. Breslauer, D. P. Rameta, A. S. A. Minetti. *J. Mol. Biol.* **365**, 1 (2007).
34. T. K. Chiu, M. Kaczor-Grzeskowiak, R. E. Dickerson. *J. Mol. Biol.* **292**, 589 (1999).
35. D. S. Goodsell, M. Kaczor-Grzeskowiak, R. E. Dickerson. *J. Mol. Biol.* **239**, 79 (1994).