Pure & Appl. Chem., Vol. 65, No. 9, pp. 1939–1946, 1993.
Printed in Great Britain.

Structural energetics of protein stability and folding cooperativity

Kenneth P. Murphy and Ernesto Freire

Department of Biology and Biocalorimetry Center, The Johns Hopkins University, Baltimore, MD 21219 USA

Abstract

Numerous studies have demonstrated that the folding/unfolding transitions of globular proteins involve very few or no thermodynamically stable intermediate structures between the folded and unfolded states. Recently we have developed a hierarchical partition function formalism aimed at gaining an understanding of the cooperative nature of thermal transitions in proteins. The energetic terms in the partition function are correlated to structural properties of the protein, namely buried surface areas and number of residues. Using phosphoglycerate kinase and myoglobin as examples, it is shown that intermediately folded states are destabilized by two features: the unfavorable exposure of apolar surface area on regions of the intermediate structure remaining folded, and a decreased gain in configurational entropy for portions of the polypeptide chain which are adjacent to those regions remaining folded.

INTRODUCTION

One of the most striking features of protein folding is its high degree of cooperativity. Cooperative interactions essentially reduce to zero the probability of hundreds of thousands of partially folded intermediates with a potential to become populated during the folding process. From a mechanistic point of view, the most fundamental questions deal with the molecular origin of this cooperative behavior. How is the probability of thermodynamically stable intermediates reduced to negligible levels? Why do single domain proteins undergo two-state transitions? How do different domains in multi-domain proteins interact with each other and define specific folding/unfolding patterns? What are the mechanisms that give rise to specific folding intermediates like molten globules? etc.

From a statistical thermodynamic standpoint, the development of a complete description of the folding/unfolding equilibrium in proteins requires the specification of the system partition function, Q, defined as the sum of the statistical weights of all the possible states of the molecule ($Q=\Sigma\exp[-\Delta G_i/RT]$). There are two central elements in calculating the folding/unfolding partition function from the protein structure. First is the identification and enumeration of all accessible states of the protein; and second is the assignment of a Gibbs free energy function to each of those states.

The ensemble of all accessible conformational states of the protein is astronomical ($\sim X^N$ where X is the number of configurations per residue and N is the number of residues in the protein). Obviously, a complete enumeration of all the states in the partition function is a computationally intractable task. A sensible approach is to include in the partition function only those terms that have the highest probabilities and to use the crystallographic structure as a template to generate this sub-ensemble.

The strategy that we have used to formulate the folding/unfolding partition function for a protein of known crystallographic structure is to define hierarchical levels of cooperative folding units. At the most fundamental level are those structural elements that exhibit a two-state behavior as a result of purely local or intrinsic interactions with respect to chain sequence, whereas at higher levels are those structural motifs that behave in this manner as a result of longer range interactions with more distant structural elements in the protein chain. α-helices inside proteins are examples of low order, and in many cases fundamental, cooperative folding units, whereas entire domains or even whole proteins may, under some conditions, behave as higher order cooperative folding units. We have used interaction specific three-dimensional contact plots calculated from crystallographic structures in order to identify cooperative folding units within proteins.

The free energy of any particular state of the protein relative to that of the native state is the sum of several contributions, primarily: the disruption of contacts and exposure of apolar surfaces to solvent water; the disruption of polar interactions (hydrogen bonds and polar van der Waals) and the exposure of the polar

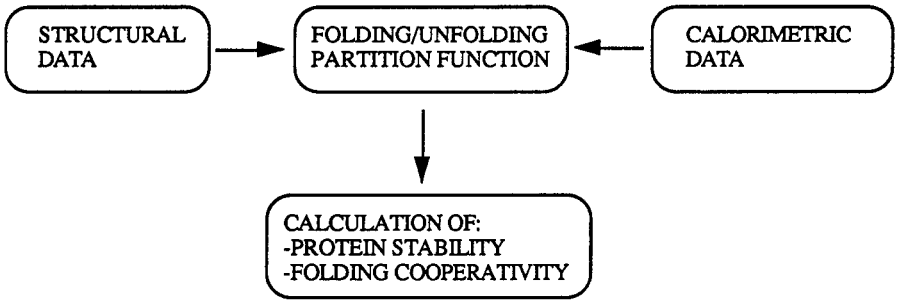


Fig. 1. Schematic of structural thermodynamic algorithms aimed at predicting protein stability and protein folding/unfolding cooperativity.

surfaces to solvent water; the change in configurational degrees of freedom of the protein; and other more specific changes including the protonation of ionizable groups, salt bridges, disulfide bonds, or changes in the number of bound ligands. In principle, the contribution of each interaction to the free energy of stabilization can be evaluated from the crystal structure in conjunction with the fundamental thermodynamic values (ΔH, ΔS, and ΔC_p) that characterize each type of interaction. Only recently, a sufficient experimental understanding of the magnitude of the individual forces that determine the stability of proteins has been gained.

The statistical thermodynamic dissection of protein structure, together with the calorimetrically derived set of fundamental thermodynamic parameters provide a solid framework for the development of structural thermodynamic algorithms aimed at predicting protein stability and protein folding/unfolding cooperativity as illustrated in Fig. 1.

METHODS

The hierarchical partition function

As mentioned above the strategy employed here is to formulate the folding/unfolding partition function for a protein of known crystallographic structure and to define hierarchical levels of cooperative folding units. The partition function, Q, for any system is simply the sum of the statistical weights of all the accessible states of the system:

$$Q = \Sigma \exp[-\Delta G_i / RT] \quad (1)$$

$$= \Sigma K_i \quad (1a)$$

where R is the gas constant, T is the absolute temperature in kelvin, and ΔG_i is the difference in free energy between state i and the reference state. Here the reference state is chosen as the folded, or native, state of the protein. As regards folding/unfolding transitions in proteins, the partition function can be written as:

$$Q = 1 + \Sigma \exp[-\Delta G_i / RT] + \exp[-\Delta G_D / RT] \quad (2)$$

$$= 1 + \Sigma K_i + K_D \quad (2a)$$

where the initial term is the statistical weight of the native state, the final term is the statistical weight of the fully denatured state, and the middle term is the sum of the statistical weights of all the intermediate species.

For small globular proteins, the surprising observation is that the partition function for the folding/unfolding transition simplifies to the two-state partition function:

$$Q = 1 + \exp[-\Delta G_D / RT] \quad (3)$$

$$= 1 + K_D \quad (3a)$$

In other words the statistical weights for all intermediate species become negligibly small. How is this cooperativity obtained?

In order to understand how the population of the intermediate states are reduced to nearly zero, the partition function can be decomposed in a hierarchical sense by considering the protein to be composed of fundamental cooperative units, and then considering the interaction between these units. This can best be understood by considering the system illustrated in Fig. 2. Here we have a protein composed of two cooperative units: **1** and **2**. The intrinsic free energy difference of each cooperative unit is designated as

ΔG_i and has a statistical weight of κ_i, while the interaction between the cooperative units gives rise to an additional free energy terms, Δg_i, having a statistical weight ϕ_i. The partition function then is written as:

$$Q = 1 + \kappa_1\phi_1 + \kappa_2\phi_2 + \kappa_3\phi_3 \qquad (4)$$

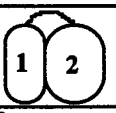
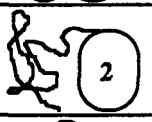
STATE	FREE ENERGY	RELATIVE FREE ENERGY	STATISTICAL WEIGHT
	$G_1 + G_2 + g_{1,2}$	0	1
	$G_{1'} + G_2 + g_{1',2}$	$\Delta G_{1'} + \Delta g_{1',2}$	$\kappa_1\phi_1$
	$G_1 + G_{2'} + g_{1,2'}$	$\Delta G_{2'} + \Delta g_{1,2'}$	$\kappa_2\phi_2$
	$G_{1'} + G_{2'} + g_{1',2'}$	$\Delta G_{1'} + \Delta G_{2'} + \Delta g_{1',2'}$	$\kappa_3\phi_3$

Fig. 2. Schematic representation of the energetic states for a protein composed of two cooperative units. (Adapted from Freire et al. (ref. 3))

It should be noted that in terms of normal equilibrium constants, $\kappa_1 = K_1$, $\kappa_2 = K_2$ and $\kappa_3 = K_1K_2$. The condition for cooperativity is that $\phi_3 \neq \phi_1\phi_2$. If $\phi_3 = \phi_1\phi_2$ then the partition function reduces to that for two independent domains:

$$Q = (1+K_1)\,(1+K_2) \qquad (5)$$

On the other hand, when ϕ_1 and ϕ_2 approach zero, then the partition function reduces to that for a two-state transition as given in Eqs. 3 and 3a.

In general, any protein can be considered to be composed of *n* cooperative units, giving rise to a partition function of 2^n terms.

$$Q = 1 + \Sigma\kappa_i\phi_i + \kappa_D\phi_D \qquad (6)$$

Furthermore, each cooperative unit can, if necessary, be considered as composed of more fundamentally cooperative elements giving rise to a nested partition function (ref. 1 - 5). The construction of the hierarchical partition function thus requires specification of cooperative units and assignment of quantitative values for the κ_i and ϕ_i terms.

In order to clarify the origins of cooperative folding/unfolding behavior, each interaction term, ϕ_i, can be further partitioned into a term arising from the exposure of surface and disruption of bonds on the cooperative unit(s) unfolded in state i, designated ϕ'_i, and a term arising from the additional surface exposed on that portion of the protein remaining folded, ϕ^*_i which is referred to as the *complementary* region.

Identification of cooperative units

In principle, cooperative units can be defined at any level, from individual amino acid residues up to entire proteins in many cases. Regardless of the level of detail required, it is of interest to correlate the assignment of cooperative units with known structural features. In order to facilitate this we have made use of contact plots. The contact plot is constructed by computing the accessible surface area (ref. 6) of each residue in a protein in the presence and absence of all its neighbors. The amount of surface area on a given residue *x* which is buried by a given residue *y* is then indicated in gray scale with no interaction being indicated as white and interactions of 25 $Å^2$ or greater being indicated as black. The contact plot can be constructed for either the apolar, polar, or total surface areas.

The apolar contact plot for phosphoglycerate kinase (PGK) is illustrated in Fig. 3. The two domains of PGK are clearly visible in the contact plot as well as the interaction of the N and C domains. The figure clearly suggests that the N and C terminal domains could be considered as cooperative units giving rise to a partition function of four states as in Eq.(4).

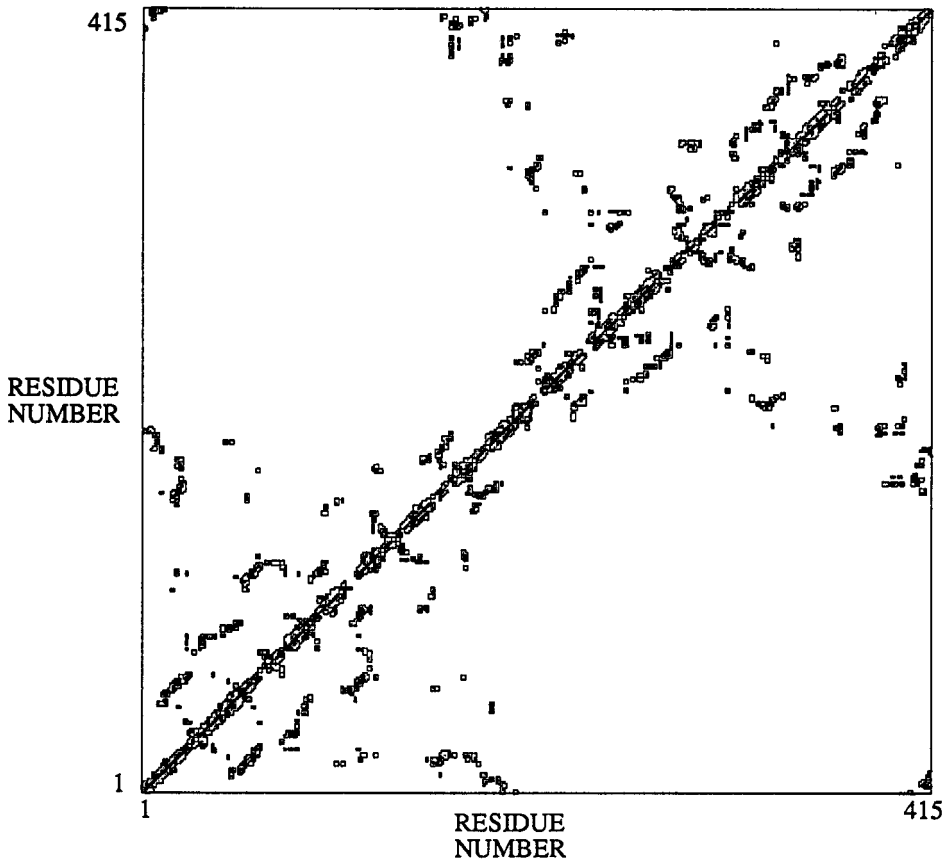


Fig. 3. Apolar contact plot for phosphoglycerate kinase.

In contrast, Fig. 4 illustrates the apolar contact plot for myoglobin. It can be seen from the figure that no clear demarcation of domains is apparent. At a more detailed level, however, the individual helices are readily identified as the thick regions of contacts along the diagonal. If the eight helices are each considered as a cooperative unit, then the partition function will have 256 (2^8) states.

Calculation of energetic parameters

In analyzing the energetics of folding/unfolding transitions in proteins, there are three primary contributions that must be considered. These are the energetics resulting from the transfer of apolar surface from the protein interior into solvent water, the transfer of apolar surface from the protein interior into solvent water, and the configurational entropy change resulting from the greater degrees of freedom which the protein chain has available to it in the denatured state. Additionally, specific contributions such as ligand binding, the presence of disulfide bonds and protonation effects must also be included as necessary.

It should be noted that the division of energetics used here differs somewhat from the standard division into the hydrophobic effect, hydrogen bonding, and configurational entropy. In our usage, the apolar contribution includes not only the restructuring of water around apolar groups exposed to solvent, but also the difference in van der Waals interaction of the apolar groups between the protein interior and water. Likewise, the polar exposure also includes not only the disruption of internal hydrogen bonds and subsequent formation of hydrogen bonds with solvent, but also any differences in van der Waals interactions.

The free energy change for protein stability can thus be written as:

$$\Delta G = \Delta G_{ap} + \Delta G_{pol} + \Delta G_{conf} + \Delta G_{other} \qquad (7)$$

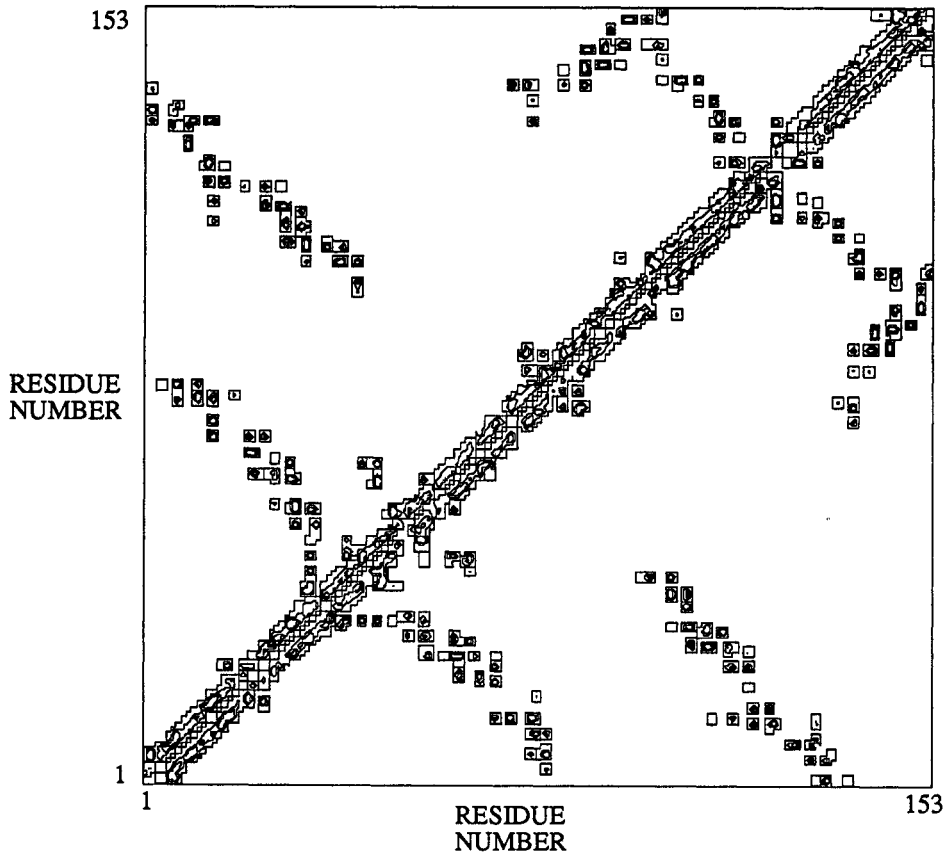


Fig. 4. Apolar contact plot for myoglobin.

where the subscripts indicate: ap - apolar surface exposure, pol - polar surface exposure, conf - configurational entropy changes, and other - case specific contributions as described above.

The free energy change for any interaction is a function of temperature according to the standard formula:

$$\Delta G = \Delta H(T_R) - T\,\Delta S(T_R) + \Delta C_p[(T-T_R) - T \ln(T/T_R)] \qquad (8)$$

where $\Delta H(T_R)$ is the enthalpy change and $\Delta S(T_R)$ is the entropy change, at the reference temperature T_R, and ΔC_p is the heat capacity change. For convenience, we choose the reference temperatures for ΔH and ΔS as those temperatures at which the apolar contribution to the respective quantity is zero. These temperatures are designated T_H^* and T_S^*. With these reference temperatures, the free energy change is written as:

$$\Delta G = \Delta H^* - T\,\Delta S^* + (\Delta C_{p,ap} + \Delta C_{p,pol})[(T-T_H^*) - T \ln(T/T_S^*)] \qquad (9)$$

where ΔH^* and ΔS^* are the enthalpy and entropy changes at T_H^* and T_S^* respectively, and the subscripts ap and pol designate the apolar and polar contributions to the heat capacity change. Since, by definition, there is no apolar contribution to the respective thermodynamic quantity at the reference temperatures ΔH^* is primarily the polar contribution to ΔH and ΔS^* the configurational contribution to the entropy change.

It has been shown that the energetics associated with the transfer of apolar or polar surface area into water is proportional to the accessible surface area (ASA) (ref. 5, 7 - 9). Analysis of model compound transfer data and the existing protein thermodynamic database provide a set of fundamental thermodynamic parameters for relating buried surface area in proteins to the terms in Eq. (9) (ref. 4, 5).

$$\Delta C_{p,ap} = 1.9\,A_{ap}\ (J\,K^{-1}\,mol^{-1}) \qquad (10)$$

$$\Delta C_{p,pol} = -1.1\,A_{pol}\ (J\,K^{-1}\,mol^{-1}) \qquad (11)$$

$$\Delta H^* = 146\,A_{pol}\ (J\,mol^{-1}) \qquad (12)$$

$$T_H^* = 373.15\ K \quad (13)$$

$$T_S^* = 385.15\ K \quad (14)$$

The only remaining term in Eq. (9) is ΔS^* which can be given as an average contribution per amino acid residue as:

$$\Delta S^* = 18\ N_{res}\ (J\ K^{-1}\ mol^{-1}) \quad (15)$$

where N_{res} is the number of amino acid residues in the protein.

The term ΔS^* can be estimated more precisely by considering the various contributions of the different side chains (ref. 10) as:

$$\Delta S^* = \Sigma\ N_i\ \Delta S_i\ (J\ K^{-1}\ mol^{-1}) \quad (16)$$

where N_i is the number of amino acids of type i and ΔS_i is the configurational entropy contribution of that amino acid type. Estimated values of ΔS_i have been calculated from the data of O'Neil and DeGrado (ref. 11) as previously described (ref. 10) and are summarized in Table 1.

TABLE 1. Estimated configurational entropy changes for individual amino acid residues upon protein unfolding. Units are $J\ K^{-1}\ mol^{-1}$. (Adapted from (ref. 4))

Residue	ΔS^*	Residue	ΔS^*
ALA	13	LEU	14
ARG	13	LYS	14
ASN	22	MET	16
ASP	21	PHE	18
CYS	20	PRO	13
GLN	18	SER	18
GLU	19	THR	22
GLY	23	TRP	17
HIS	22	TYR	21
ILE	20	VAL	21

Equations (9) - (16) have been used to estimate the folding/unfolding transition energetics of many globular proteins and it is found that the calculated values of ΔH, ΔS, and ΔC_p agree with the experimentally determined values nearly to within experimental error (ref. 4, 5).

APPLICATION TO PHOSPHOGLYCERATE KINASE

The procedures outlined above, along with Eqs. (9) - (16), allow for specification of the hierarchical partition function. As an example, consider the protein phosphoglycerate kinase (PGK). PGK is a globular protein of 415 residues with two clear structural domains (N and C) with a largely hydrophobic interface between them (ref. 12). The reversible folding/unfolding of PGK has been studied in GuHCl solutions (ref. 3, 13). The striking observation is that the thermal transition shows one peak in for the high-temperature (heat denaturation) transition at 40°C but two peaks in the low-temperature (cold denaturation) transition at 20°C indicating a large change in the cooperative free energy over a small temperature range.

If each of the two domains is considered as a cooperative unit, then the partition function takes on the form:

$$Q = 1 + K_1\ \phi'_1\ \phi^*_1 + K_2\ \phi'_2\ \phi^*_2 + K_1\ K_2\ \phi'_1\ \phi'_2 \quad (17)$$

The interface between the two domains buries 1385 $Å^2$ of apolar surface area (726 $Å^2$ on the N domain and 659 $Å^2$ on the C domain) and contains nine hydrogen bonds (ref. 5). In order to model the structural energetics of PGK under the experimental conditions, the effect of GuHCl must also be included. This has

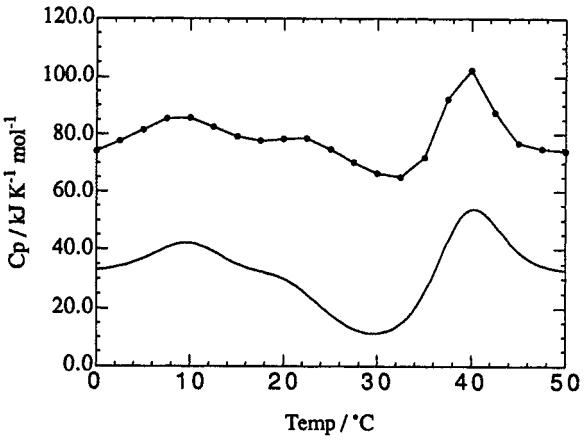


Fig. 5. Experimental (circles) (ref. 15) and theoretical (bottom curve) heat capacity functions for the thermal folding/unfolding transition of PGK at pH 6.5 in the presence of 0.7 M GuHCl. The curves have been shifted in the vertical axis for display purposes. (Taken from (ref. 3))

been done using the site binding model of Tanford (ref. 14) as previously described (ref. 3). Figure 5 shows the calculated and experimental excess heat capacity plots for PGK in 0.7M GuHCl. It can be seen that the structural energetic calculations correctly predict the change in cooperativity between the heat and cold denaturation processes.

The origins of this change in cooperativity over the 20°C are graphically illustrated in Fig. 6. In this figure the overall ΔG is shown as a function of temperature along with the free energy associated with the complementary surface exposed on either of the domains when the other is unfolded, ΔG^*. The overall ΔG crosses zero at two points corresponding to the heat and cold denaturation temperatures. It can be seen that at the heat denaturation temperature the ΔG^* terms are large and positive, while at the cold denaturation temperature they are nearly zero.

The ΔG^* terms arise from the exposure of the apolar complementary surface previously buried by the domain which is unfolded. While the unfolding domain is also exposing apolar surface, the unfavorable free energy associated with this is compensated for by the configurational entropy change gained upon unfolding. In contrast, there is no significant corresponding gain in configurational entropy for the complementary surface exposed on the remainder of the protein. This uncompensated complementary free energy is a primary source of cooperative folding/unfolding behavior.

APPLICATION TO MYOGLOBIN

The hierarchical partition function formalism can be extended to greater detail than illustrated in the above case of PGK. Figure 4 shows the apolar contact plot for myoglobin. As mentioned above, the eight helices, generally referred to as A-H, can each be considered as cooperative folding units giving rise to 256 states in the partition function. The terms in the partition function are determined from the crystallographic structure by analyzing the polar and apolar surfaces buried within each helix and between each pair of helices (ref. 2). This analysis leads to the excess heat capacity function in Fig. 7, which closely matches the experimentally determined function.

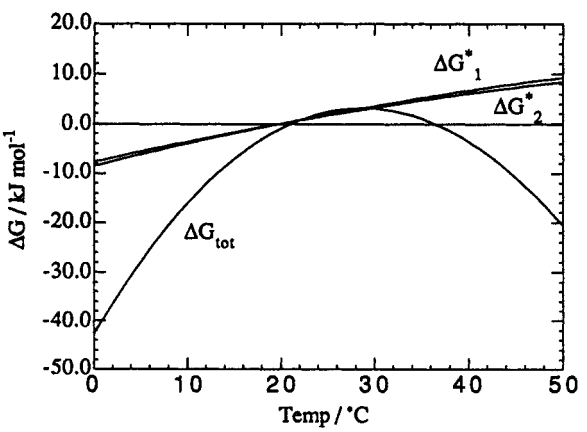


Fig 6. Calculated overall free energy of stabilization (ΔG_{total}) for PGK under the same conditions as Fig. 5. The curve displays two zeros, corresponding to the temperatures of cold and heat denaturation. Also shown are the cooperative Gibbs free energies (ΔG^*) associated with the uncompensated exposure of apolar surfaces upon unfolding of each of the domains. (Taken from (ref. 3))

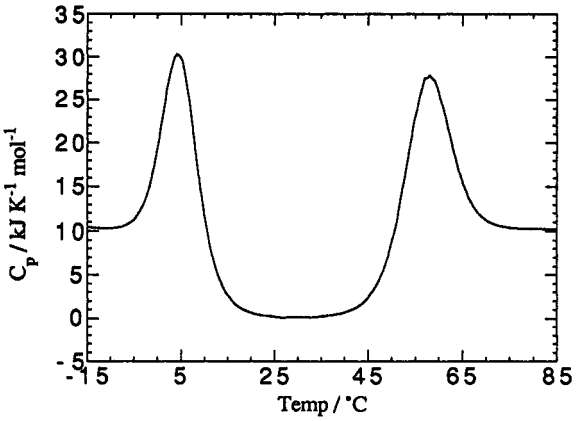


Fig. 7. Predicted excess heat capacity function versus temperature for myoglobin. The curve simulates the experimental curve obtained at pH 3.83 by Privalov *et al.* (ref. 16). Under these conditions both the cold and heat denaturation curves can be studied experimentally. The predicted values are $T_{m,cold}$ = 277 K; $T_{m,heat}$ = 331 K; ΔH = 247 kJ mol^{-1}; ΔC_p = 10.3 kJ K^{-1} mol^{-1}. The experimental values are $T_{m,cold}$ = 276 K; $T_{m,heat}$ = 330.7 K; ΔH = 222 kJ mol^{-1}; ΔC_p = 10.5 kJ K^{-1} mol^{-1} (ref. 16). (Taken from (ref. 2)).

Analysis of the partition function indicates that the sum total population of the 254 intermediate states is never greater than 10^{-6} in the transition region (ref. 2) indicating that the transition is very well represented as two-state. This cooperativity arises, to a large extent, from the uncompensated complementary surfaces exposed in each of the intermediate states. It is only in the fully unfolded state that there is a corresponding gain in configurational entropy for all the apolar surface exposed. Additionally, cooperativity also arises from the fact that when a helix within the protein unfolds while the rest of the protein remains intact, the configurational entropy gained by those residues is less than when that region of the protein gains when the entire protein unfolds (ref. 2). This results from the necessity of keeping the ends of the chain fixed to the folded portion of the protein. The additional configurational entropy gained upon complete unfolding constitutes an extra source of cooperative free energy.

CONCLUSIONS

The hierarchical partition function analysis discussed here represents a novel approach to connecting structural and energetic features of proteins and utilizing these structural energetic features to address a specific problem in protein folding; namely, why do globular proteins undergo highly cooperative folding/unfolding transitions. The examples illustrate that cooperativity has two primary sources. These are the uncompensated exposure of the complementary surface which accompanies all folding intermediates but is absent in the fully folded and unfolded states, and the additional configurational entropy gained by the unfolding of a region of a protein when the ends of that region are not fixed.

Acknowledgement

Supported by grants from the National Institutes of Health RR04328, GM37911 and NS24520.

REFERENCES

1. C.H. Robert, H. Decker, B. Richey, S.J. Gill and J. Wyman, Proc. Natl. Acad. Sci. USA 84, 1891-1895 (1987).
2. E. Freire and K.P. Murphy, J. Mol. Biol. 222, 687-698 (1991).
3. E. Freire, K.P. Murphy, J.M. Sanchez-Ruiz, M.L. Galisteo and P.L. Privalov, Biochemistry 31, 250-256 (1992).
4. K.P. Murphy, V. Bhakuni, D. Xie and E. Freire, J. Mol. Biol. 227, 293-306 (1992).
5. K.P. Murphy and E. Freire, Adv. Protein Chem. 43, 313-361 (1992).
6. B. Lee and F.M. Richards, J. Mol. Biol. 55, 379-400 (1971).
7. R.B. Hermann, J. Phys. Chem. 76, 2754-2759 (1972).
8. D. Eisenberg and A.D. McLachlan, Nature 319, 199-203 (1986).
9. R.S. Spolar, J.-H. Ha and M.T. Record Jr., Proc. Natl. Acad. Sci. USA 86, 8382-8385 (1989).
10. K.P. Murphy, D. Xie, K.C. Garcia, L.M. Amzel and E. Freire, Proteins 15, 1 (1993).
11. K. O'Neil and W. DeGrado, Science 250, 646-651 (1990).
12. H.C. Watson, N.P.C. Walker, P.J. Shaw, P.L. Bryant, L.A. Fothergill, R.E. Perkins, S.C. Conroy, M.J. Dobson, M.F. Tuite, A.J. Kingsman and S.M. Kingsman, EMBO J. 1, 1635-1640 (1982).
13. Y.V. Griko, S.Y. Venyaminov and P.L. Privalov, FEBS Lett. 244, 276-278 (1989).
14. K.C. Aune and C. Tanford, Biochemistry 8, 4579-4590 (1969).
15. Y.V. Griko, S.Y. Venyaminov and P.L. Privalov, FEBS Lett. 244, 276-278 (1989).
16. P.L. Privalov, Y.V. Griko, S.Y. Venyaminov and V.P. Kutyshenko, J. Mol. Biol. 190, 487-498 (1986).