The Molecular Basis of Cooperativity in Protein Folding. Thermodynamic Dissection of Interdomain Interactions in Phosphoglycerate Kinase†

Ernesto Freire,*,‡ Kenneth P. Murphy,‡ Jose M. Sanchez-Ruiz,§ Maria L. Galisteo,§ and Peter L. Privalov∥

Department of Biology and Biocalorimetry Center, The Johns Hopkins University, Baltimore, Maryland 21218, Department of Physical Chemistry, University of Granada, Granada, Spain, and Institute of Protein Research, Academy of Sciences of the USSR, Pushchino, Moscow Region, USSR

Received May 17, 1991; Revised Manuscript Received August 15, 1991

ABSTRACT: In the presence of guanidine hydrochloride, phosphoglycerate kinase from yeast can be reversibly denatured by either heating or cooling the protein solution above or below room temperature [Griko, Y. V., Venyaminov, S. Y., & Privalov, P. L. (1989) *FEBS Lett. 244*, 276–278]. The heat denaturation of PGK is characterized by the presence of a single peak in the excess heat capacity function obtained by differential scanning calorimetry. The transition curve approaches the two-state mechanism, indicating that the two domains of the molecule display strong cooperative interactions and that partially folded intermediates are not largely populated during the transition. On the contrary, the cold denaturation is characterized by the presence of two peaks in the heat capacity function. Analysis of the data indicates that at low temperatures the two domains behave independently of each other. The crystallographic structure of PGK has been used to identify the nature of the interactions between the two domains. These interactions involve primarily the apposition of two hydrophobic surfaces of approximately 480 Å^2 and nine hydrogen bonds. This information, in conjunction with experimental thermodynamic values for hydrophobic, hydrogen bonding interactions and statistical thermodynamic analysis, has been used to quantitatively account for the folding/unfolding behavior of PGK. It is shown that this type of analysis accurately predicts the cooperative behavior of the folding/unfolding transition and its dependence on GuHCl concentration.

The process of protein folding is characterized by a high degree of cooperativity. This cooperativity is expressed in the almost negligible population levels of partially folded stable intermediate states observed during folding/unfolding transitions. For example, for most small globular proteins (M_r <20000), the population of partially folded intermediates never exceeds 5% (Privalov & Khechinachvilli, 1974; Freire & Biltonen, 1978; Privalov, 1979). Folding/unfolding transitions of multidomain proteins, on the other hand, are often characterized by the presence of a few partially folded equilibrium intermediates that usually correspond to conformations in which some domains are folded and other ones unfolded (Privalov, 1982; Ramsay et al., 1989; Brandts et al., 1989; Ramsay & Freire, 1990). The population of these stable equilibrium intermediates is determined by the magnitude of the interdomain interactions.

The elucidation of the physical mechanisms that generate the cooperative folding behavior of proteins is currently the focus of intense research. Recently, Freire and Murphy (1991) developed a hierarchical algorithm aimed at calculating the folding/unfolding partition function of proteins and estimating the population of structurally defined partially folded intermediates. This algorithm considers a protein as being formed by hierarchical layers of interacting cooperative folding units. Cooperative folding units are defined as structural elements whose folding/unfolding behavior can be described as two-state processes. Within this context, first-order cooperative folding units are those structural elements that behave as such due to local interactions. Higher order cooperative folding units are constructed from lower order ones through longer range interactions, i.e., through interactions involving amino acids in different structural elements. Small globular proteins exhibiting two-state behavior are examples of higher order cooperative folding units. Individual domains in multidomain proteins also behave in many cases as higher order cooperative folding units (Privalov, 1982; Brandts et al., 1989; Ramsay & Freire, 1990).

Recently, Griko et al. (1989) observed that in the presence of guanidine hydrochloride, the heat denaturation of the two-domain protein phosphoglycerate kinase (PGK) from yeast is a highly cooperative two-state process whereas the cold denaturation is characterized by two separate transitions corresponding to the unfolding of the two structural domains of the protein. Since the crystallographic structure of PGK is known, this protein appears to be an ideal candidate to perform free energy calculations of the domain interface in an attempt to account for its cooperative folding/unfolding behavior.

Theory

The 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 (Freire & Bitonen, 1978):

$$Q = \sum_{i=0}^{n} e^{-\Delta G_i/RT} \quad (1)$$

where ΔG_i is the Gibbs free energy difference between the ith state and the reference state, R the gas constant, and T the absolute temperature. Following the standard convention used

†Supported by Grants RR04328 and NS24520 from the National Institutes of Health (to E.F.) and by Grant PB87-0871 from the DGICYT (to J.M.S.). K.P.M. was supported by a postdoctoral fellowship from the National Science Foundation (DIR-8721059).

*To whom correspondence should be addressed.
‡The Johns Hopkins University.
§University of Granada.
∥Academy of Sciences of the USSR.

0006-2960/92/0431-250$03.00/0

in protein thermodynamics, the reference state is taken as the native state (Privalov, 1979). For convenience, eq 1 can be written as

$$Q = 1 + \sum_{i=1}^{n-1} e^{-\Delta G_i/RT} + e^{-\Delta G_n/RT} \quad (2)$$

where the first and last terms on the right-hand side are the statistical weights of the native and unfolded states, respectively, and the summation represents the statistical weights of all the partially folded intermediate states.

In theory, the total number of partially folded states is astronomical; in practice, however, the number of partially folded intermediates that ever become populated is very small. For many small globular proteins, it has been shown that the population of intermediates never exceeds 5% and that the folding/unfolding transition can be well approximated by the two-state partition function $Q = 1 + e^{-\Delta G_n/RT}$ (Freire & Biltonen, 1978; Privalov, 1979). Free energy calculations for partially folded intermediates of myoglobin (Freire & Murphy, 1991) are also consistent with this behavior and demonstrate that the crystallographic structure of a protein can be used to predict its folding/unfolding behavior.

For multidomain proteins, on the other hand, the folding/unfolding transition is often characterized by the presence of a few intermediates that usually but not necessarily correspond to states in which some structural domains are folded and other ones unfolded (Privalov, 1982; Brandts et al., 1989; Ramsay & Freire, 1990). The extent to which these partially folded states become populated depends on the magnitude of the interactions between domains. Several forces are involved in these interactions. They include hydrophobic interactions, hydrogen bonds, and van der Waals and electrostatic interactions as well as ligand binding effects to name a few. Crystallographic information in conjunction with experimental thermodynamic values for these interactions and statistical thermodynamic analysis should account for and predict the folding/unfolding behavior.

Thermodynamic Dissection of Cooperative Interactions. The forces that stabilize the structure of proteins give rise to a well-defined set of cooperative folding/unfolding patterns. These patterns will be illustrated for the case of a protein consisting of two cooperative folding units. As shown in Figure 1, the number of states associated with the folding/unfolding behavior of this hypothetical protein correspond to the total number of combinations generated by folding and unfolding its constituent folding units. These states correspond to the folded structure, two partially folded intermediate states, and the unfolded state (Ramsay & Freire, 1990; Freire & Murphy, 1991). The partition function of such a molecule can be written as

$$Q = 1 + \Phi_1 K_1 + \Phi_2 K_2 + \Phi_3 K_3 \quad (3a)$$

$$Q = 1 + \Phi_1 K_1 + \Phi_2 K_2 + \Phi_3 K_1 K_2 \quad (3b)$$

where the $K_i = e^{-\Delta G_i/RT}$ terms represent the intrinsic statistical weights and the $\Phi_i = e^{-\Delta g_i/RT}$ terms represent the interaction parameters. The K_i terms contain the contributions to the structural stability that originate from interactions within each folding unit, whereas the Φ_i terms contain the contributions arising from interactions between folding units. In eq 3a, all the free energy terms are referenced to the initial state as detailed in Figure 1. The partition function can also be written in terms of the intrinsic stability constants for the first (K_1) and second (K_2) domains as shown in eq 3b. Previously, Brandts et al. (1989) have presented a model in which only interactions in the folded state are considered; as shown before,

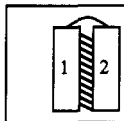
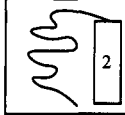
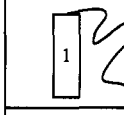
STATE	FREE ENERGY	RELATIVE FREE ENERGY	STAT. WEIGHT
	$G_1 + G_2 + g_{1,2}$	0	1
	$G_{1'} + G_2 + g_{1',2}$	$\Delta G_{1'} + \Delta g_{1',2}$	$\Phi_1 K_1$
	$G_1 + G_{2'} + g_{1,2'}$	$\Delta G_{2'} + \Delta g_{1,2'}$	$\Phi_2 K_2$
	$G_{1'} + G_{2'} + g_{1',2'}$	$\Delta G_{1'} + \Delta G_{2'} + \Delta g_{1',2'}$	$\Phi_3 K_3$

FIGURE 1: Schematic representation of states, free energy assignments, and statistical weights for a protein composed of two cooperative folding units. The folding/unfolding partition function is defined as the sum of the statistical weights.

that model represents a particular case of eq 3 (Ramsay & Freire, 1990). From a purely mathematical point of view, cooperativity is generated by the condition $\Phi_3 \neq \Phi_1\Phi_2$. If $\Phi_3 = \Phi_1\Phi_2$, then the partition function reduces to

$$Q = 1 + \Phi_1 K_1 + \Phi_2 K_2 + \Phi_1\Phi_2 K_1 K_2 \quad (4a)$$

$$Q = (1 + \Phi_1 K_1)(1 + \Phi_2 K_2) \quad (4b)$$

which is mathematically identical to the partition function for two independent domains.

The cooperative behavior of the system is determined by the properties of the Φ_i terms. For convenience, these terms can be expressed as the product of two terms ($\Phi_i = \Phi_i^* \Phi_i'$) where the Φ_i' values contain the contributions from the protein region(s) undergoing unfolding and the Φ_i^* terms contain the contributions from the remaining regions of the protein (here termed complementary regions) (Freire & Murphy, 1991). This operational definition allows a mechanistic dissection of the different types of interactions that contribute to the cooperative folding/unfolding behavior of a protein. The partition function for the protein in Figure 1 is then written as

$$Q = 1 + \Phi_1^* \Phi_1' K_1 + \Phi_2^* \Phi_2' K_2 + \Phi_3' K_3 \quad (5)$$

or in general

$$Q = 1 + \sum_{i=1}^{n-1} \Phi_i^* \Phi_i' K_i + \Phi_n K_n \quad (6)$$

It must be noted that, unlike the partially folded states, the totally unfolded state lacks a complementary region. The immediate consequence of this property is that partially folded states always expose to the solvent structural regions for which there is not a corresponding configurational entropy gain (Freire & Murphy, 1991).

Classification of Cooperative Mechanisms. (a) Hydrophobic Interface. The interactions between domains or subdomains in proteins often involve the close apposition of hydrophobic surfaces. As illustrated in Figure 2A, the unfolding of one of the domains elicits the exposure of apolar residues in the domain undergoing unfolding as well as the rest of the protein. While the solvent exposure of the unfolded domain is compensated by the favorable free energy of unfolding and especially its positive conformational entropy, the solvent exposure of the complementary regions of the protein is not compensated by a similar entropy change. Energetically, all

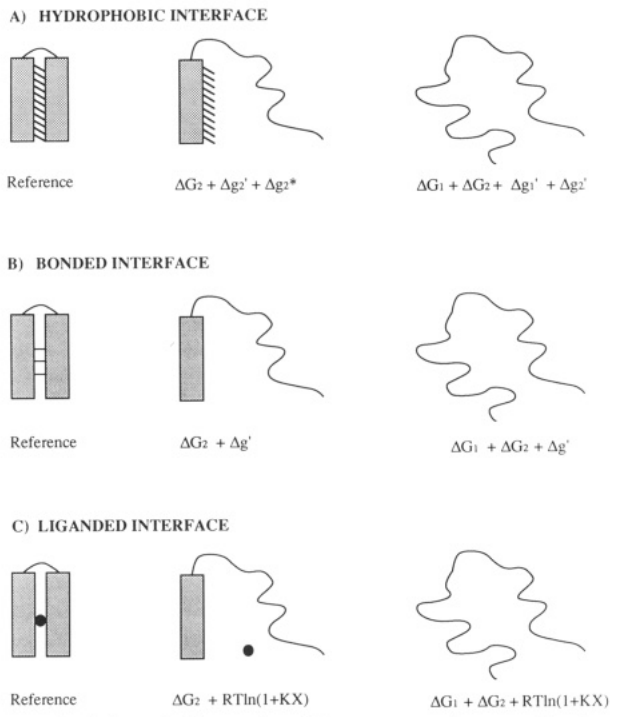


FIGURE 2: Schematic illustration of free energy assignments for three different types of interactions. The examples in the figure illustrate the situation existing for the folded, unfolded, and one partially folded intermediate. For hydrophobic interfaces, partially folded intermediates always include an extra term Δg^* corresponding to the solvent exposure of protein regions that have not undergone unfolding. As shown in the figure, the unfolded state lacks this uncompensated exposure term. For a bonded or liganded interface, cooperative behavior is created when the unfolding of either domain results in the disruption of the bonded interface or the dissociation of the ligand molecule.

partially folded states carry an extra free energy term arising from this uncompensated exposure of apolar groups. This uncompensated free energy defines the Φ^* terms in the partition function and constitutes the source of cooperativity for the hydrophobic interaction. It can be mathematically demonstrated that for a purely hydrophobic interface, the partition function would reduce to that of independent folding units if there were no exposure of complementary regions (eq 4).

The Gibbs free energy associated with the solvent exposure of apolar regions of the protein, ΔG_{ap}, can be calculated from the total area of the apolar surface exposed to water upon unfolding or from the number of apolar hydrogens (i.e., hydrogens bonded to carbon) exposed to the solvent upon unfolding (Murphy & Gill, 1990, 1991; Freire & Murphy, 1991). These two quantities are directly proportional to each other as demonstrated before (Murphy & Gill, 1991) and as seen in calculations performed with the *n*-alkanes C1–C5. The relationship found between the surface area calculated using Connolly's algorithm (Connolly, 1983) and the number of apolar hydrogens is linear with a slope equivalent to 9.75 $Å^2$/(apolar hydrogen).

Each apolar hydrogen that becomes exposed to water contributes a $\Delta C_{p,ah}$ increment of 6.69 cal/(K·mol) (Murphy & Gill, 1990, 1991) or equivalently 0.686 cal/(K·mol) per $Å^2$ of apolar area. This change in heat capacity can then be used to calculate ΔH_{ap} and ΔS_{ap} at any temperature by using the standard thermodynamic equations:

$$\Delta H_{ap} = n_{ah}\Delta C_{p,ah}(T - T_H^*) \quad (7)$$

$$\Delta S_{ap} = n_{ah}\Delta C_{p,ah} \ln (T/T_S^*) \quad (8)$$

where n_{ah} is the number of apolar hydrogens that become exposed to water upon unfolding. T_H^* and T_S^* are the temperatures at which the ΔH_{ap} and ΔS_{ap} are zero, respectively. The value of T_S^* is well characterized for apolar transfer processes as 112 °C (Baldwin, 1986; Murphy et al., 1990). The value of T_H^* is not as well characterized but is close to 100 °C for proteins (Murphy, 1990; Murphy & Gill, 1991). As discussed before (Murphy & Gill, 1991; Freire & Murphy, 1991), ΔG_{ap} is positive at temperatures higher than 20 °C and negative at lower temperatures. The immediate prediction is that cooperative behavior based on purely hydrophobic interactions will be stronger at high temperatures and diminish at lower temperatures.

In general, the hydrophobic contribution to the cooperative behavior is dictated by the magnitude of the $\Phi_i^* = e^{-\Delta G_{ap,i}/RT}$ terms in the partition function. The magnitude of these terms is determined by the free energy change ($\Delta G_{ap,i}$) associated with the exposure to the solvent of the apolar surface on the complementary region of the protein. Since the number of apolar hydrogens can be obtained from structural data, the above equations allow calculations of ΔG_{ap} for any structurally defined state of the protein.

(*b*) *Bonded Interface.* The interaction between domains or subdomains often involves hydrogen bonds, van der Waals contacts, salt bridges, and other noncovalent interactions. This situation is illustrated in Figure 2B, where the change in Gibbs free energy associated with the unfolding of one or two domains is shown. It must be noted that the unfolding of one of the domains is sufficient to break the bonds at the interface. This situation assures a nonlinear increase of the free energy upon domain unfolding, thus triggering the cooperative behavior.

The enthalpy change for the overall process associated with the rupture of hydrogen bonds in proteins has been estimated from protein and solid model compound data (Privalov & Gill, 1988; Murphy & Gill, 1990, 1991) and is given by the equation:

$$\Delta H_{hb} = \Delta H^\circ_{hb} + \Delta C_{p,hb}(T - T_H^*) \quad (9)$$

where $T_H^* \cong 100$ °C, the reference temperature, is the temperature at which the apolar contribution to the enthalpy change is zero. At this temperature, $\Delta H_{hb} = \Delta H^\circ_{hb}$ and is estimated to be in the range of 1.1–1.8 kcal/mol from existing literature data (Privalov & Gill, 1988; Scholtz et al., 1991). $\Delta C_{p,hb}$ is equal to -14.3 ± 1.4 cal/(K·mol) (Murphy & Gill, 1990, 1991). The entropy change associated with breaking one hydrogen bond depends to a large extent on the conformational entropy change associated with the structural region affected by the event and should be evaluated individually. On the average, however, the entropy change at 112 °C is on the order of 5.5 cal/(K·mol) for an α-helix to coil transition and on the order of 4.3 cal/(K·mol) per residue for the unfolding of globular proteins [see Freire and Murphy (1991) for a discussion].

(*c*) *Liganded Interface.* As shown in Figure 2C, the energetics of a liganded interface are mathematically very similar to those of a bonded interface except for the dependence on ligand concentration. In this case also, the ligand molecule dissociates after one of the domains unfolds, giving rise to the nonlinear dependence of the Gibbs free energy. It must be noted that a liganded interface contributes to the cooperative behavior if and only if the ligand dissociates upon unfolding.

(*d*) *Other.* The interactions listed above are common to many protein systems and yield characteristic contributions to the free energy of interaction. They do not represent an exhaustive list. For example, in some cases one of the domains at the interface may contain groups that become protonated

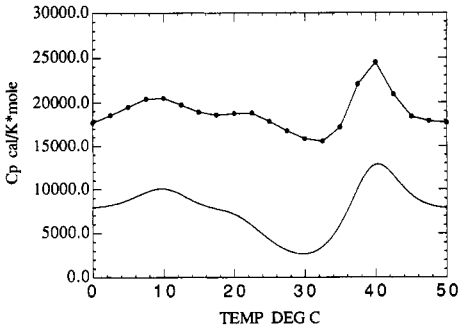


FIGURE 3: Experimental (circles) and theoretical heat capacity functions for the thermal folding/unfolding transition of phosphoglycerate kinase at pH 6.5 in the presence of 0.7 M GuHCl. The heat denaturation transition is characterized by a single peak, whereas the cold denaturation displays two peaks corresponding to the independent unfolding of the N and C domains. The experimental curve has been published before (Griko et al., 1989). As discussed in the text, the theoretical curve does not represent the best fit to the experimental data, but the calculated curve using structural information in conjunction with thermodynamic information for elementary interactions. The curves have been shifted in the vertical axis for display purposes.

upon unfolding. In this case, protonation occurs either upon unfolding of the structural domain that contains the ionizable groups or upon unfolding of the corresponding complementary region. The corresponding free energy must be included in the Φ terms of all the states that expose the ionizable group. This can be done if the structure of the protein is known as previously demonstrated for the case of myoglobin (Freire & Murphy, 1991).

Finally, the cooperative effects listed above are not exclusive of each other. In fact, it is perfectly feasible for two or more of them to exist simultaneously. In such cases, all the free energy contributions must be included in the appropriate Φ terms.

Results and Discussion

Differential Scanning Calorimetry. Phosphoglycerate kinase is a protein composed of two structural domains of approximately equal molecular weight (N-terminal domain M_r = 20814, and C-terminal domain M_r = 23735). The crystallographic structures of both the horse and the yeast enzymes have been determined (Banks et al., 1979; Watson et al., 1982). These structures can be used to calculate the hydrophobic surface exposed to water upon total or partial unfolding as well as the number of hydrogen bonds and other bonds involved in the stabilization of the folded conformation. For the studies presented in this paper, the structure of the yeast enzyme was used since calorimetric data are available for this molecule.

The thermal unfolding of yeast phosphoglycerate kinase has been studied before by differential scanning calorimetry in aqueous solution, where the process is irreversible (Hu & Sturtevant, 1987; Brandts et al., 1989; Galisteo et al., 1991). In the presence of increasing concentrations of GuHCl, however, the transition becomes reversible, and as such, it can be analyzed using thermodynamic methods (Griko et al., 1989). At GuHCl concentrations higher than 0.5 M, the cold denaturation as well as the heat denaturation transitions are clearly visible in the calorimetric scans (Griko et al., 1989). Analysis of the calorimetric data indicates that the heat denaturation transition approaches a two-state transition whereas the cold denaturation is characterized by the presence of two peaks, indicating that the two domains undergo separate transitions. As shown in Figure 3, at 0.7 M GuHCl the heat denaturation transition is centered at ~40 °C and characterized by an enthalpy change of 140 kcal/mol and a ΔC_p of 7.3 cal/(K·mol). The cold denaturation peaks are centered at about 7 and 20 °C. The dramatic change in cooperative behavior for the heat and cold denaturation transitions within a single calorimetric scan occurs in a temperature interval of less than 20 °C, indicating that the Φ terms in the partition function are strongly temperature-dependent. These results suggested that the primary interaction between domains was of a hydrophobic nature (Griko et al., 1989), consistent with the crystallographic structure.

Structural Energetics. As discussed before, the apolar surface that becomes exposed to the solvent upon unfolding or partial unfolding is proportional to the number of apolar hydrogens exposed to the solvent and can be calculated from the crystallographic structure of the protein (Freire & Murphy, 1991). For phosphoglycerate kinase, our calculations were directed at evaluating (a) the total number of hydrogen bonds in the N and C domains, (b) the apolar surfaces exposed by the N and C domains upon unfolding, (c) the number of hydrogen bonds at the interface between the N and C domains, and (d) the apolar surface areas contributed by the N and C domains to the interface.

The structure of the yeast enzyme was analyzed using the molecular graphics package Insight II. The apolar surface areas of both domains were determined using the Connolly algorithm (1983a,b) and converted to units of apolar hydrogens as described above. Hydrogen bonds within each domain and at the domain interface were counted directly using Insight II standard distance parameters. Analysis of the crystallographic structure indicates that upon complete unfolding of the PGK molecule a total of 1828 apolar hydrogens (814 from the N domain and 1014 from the C domain) become exposed to the solvent. Also, 75% of the polar surfaces are buried in the folded state (Chothia, 1976) and contribute −14.3 cal/(K·mol) of hydrogen bond to the heat capacity change upon unfolding (Murphy & Gill, 1990, 1991). These values predict a ΔC_p upon unfolding of 7.78 kcal/(K·mol) of PGK [3.43 kcal/(K·mol) contributed by the N domain and 4.35 kcal/(K·mol) by the C domain]. The experimental value obtained from the temperature dependence of the enthalpy change under conditions in which PGK undergoes reversible thermal unfolding is 7.5 kcal/(K·mol). The enthalpy change extrapolated to the reference temperature T_H^* = 100 °C is 564 kcal/mol. This number is consistent with an average value of 1.36 kcal/mol per amino acid residue, very close to the value of 1.35 ± 0.02 kcal/mol obtained for 12 proteins (Privalov & Gill, 1988).

Analysis of the structure also reveals that the two domains interact primarily through hydrophobic and hydrogen bond interactions at the interface. The number of apolar hydrogens that become exposed on the C domain upon unfolding of the N domain is 49.4, and the number of apolar hydrogens exposed on the N domain upon unfolding of the C domain is 44.8. These values correspond to 482 and 437 Å^2 of exposed apolar surface, respectively. In addition, nine hydrogen bonds from the overlapping carboxy and amino terminals are broken following the unfolding of the N or C domain.

The structural calculations for the interdomain interaction predict a ΔH_{ap} of −24.8 kcal/mol at 25 °C for the uncompensated exposure generated by the unfolding of the N domain and −22.5 kcal/mol for the uncompensated exposure generated by the unfolding of the C domain. The corresponding ΔS_{ap} values at 25 °C are −84.6 and −76.7 cal/(K·mol), and the ΔC_p values are 330.5 and 299.7 cal/(K·mol), respectively. These thermodynamic values completely specify the apolar contribution to the Φ^* and Φ' terms in the partition function, al-

Table I: Thermodynamic Parameters for the Thermal Unfolding of Phosphoglycerate Kinase[a]

term	ΔH (cal/mol)	ΔS [cal/(K·mol)]	ΔC_p [cal/(K·mol)]
K_1^0	2373	8.54	3130.3
K_2^0	17294	37.6	4019.5
Φ_1^*	-24786	-84.61	330.5
Φ_2^*	-22477	-76.73	299.7
Φ_1'	-22477	-76.73	299.7
Φ_2'	-24786	-84.61	330.5
Φ_{HB}	20335	50.51	-128.7

[a] The thermodynamic parameters in the table correspond to 25 °C and describe the transition observed in 20 mM phosphate buffer, pH 6.5. The effect of GuHCl is computed by means of eq 11 as explained in the text. The subscripts 1 and 2 refer to the N and C domains, respectively. As explained in the text, for a two-domain protein, $\Phi_1^* = \Phi_2'$ and $\Phi_2^* = \Phi_1'$.

lowing us to predict the cooperative behavior of the folding/unfolding transition and to compare the results with the experimental data (Freire & Murphy, 1991). For PGK, the Φ^* terms contain only uncompensated hydrophobic contributions, while the Φ' terms contain the contributions from the hydrogen bonds at the interface as well as the hydrophobic exposure of the domain undergoing unfolding. According to these definitions, the folding/unfolding partition function for PGK can be written as

$$Q = 1 + \Phi_1^*\Phi_1'\Phi_{HB}'K_1 + \Phi_2^*\Phi_2'\Phi_{HB}'K_2 + \Phi_1'\Phi_2'\Phi_{HB}'K_1K_2 \quad (10)$$

where, for clarity, the hydrophobic and hydrogen bonding contributions to the Φ' terms have been written separately. It must be noted that for a two-domain protein like PGK, $\Phi_1^* = \Phi_2'$ and $\Phi_2^* = \Phi_1'$. Table I summarizes the thermodynamic contributions of each term in the partition function.

Figure 3 shows the heat capacity function calculated with the parameters and the formalism presented in this paper. As seen in the figure, the theoretical curve accurately reproduces the features observed experimentally. The heat denaturation curve centered at 40 °C is close to a two-state transition characterized by a van't Hoff to calorimetric enthalpy ratio near unity, while the cold denaturation is characterized by two peaks corresponding to two independent transitions. It must be noted that the theoretical curve in Figure 3 does not represent the result of performing a nonlinear least-squares fitting of the data. The only parameters adjusted were the intrinsic entropy changes for the N and C domains in order to match the transition temperature of the heat denaturation peak. This adjustment is necessary to accurately locate the overall transition temperature since the existing entropy values from the protein database are not precise enough to narrowly define transition temperatures. Nevertheless, it should be mentioned that the adjusted values are on the order of 4.5 cal/(K·mol) of amino acid residue at T_S^* (as deduced from Table I by calculating the entropies at 112 °C and renormalizing them on a per mole of amino acid basis). Those values are within the range 4.3 ± 0.3 cal/(K·mol) from the protein database at that same temperature (Privalov & Gill, 1988). At T_S^*, the unfolding entropies are somewhat larger than the intrinsic entropy associated with the disruption of the hydrogen bonds at the domain interface [~2 cal/(K·mol)], perhaps reflecting the fact that this latter process does not intrinsically involve unfolding.

Figure 4 shows the overall free energy of stabilization of the N and C domains of PGK at 0.7 M GuHCl as well as the cooperative Gibbs free energies ΔG^*_1 and ΔG^*_2 corresponding

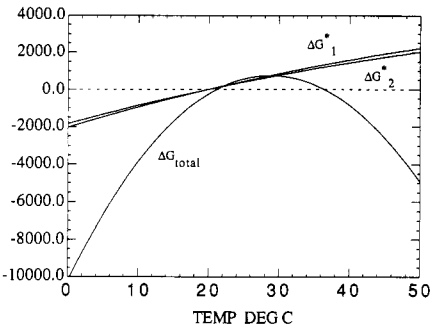


FIGURE 4: Calculated overall free energy of stabilization (ΔG_{total}) for yeast phosphoglycerate kinase at pH 6.5 and 0.7 M GuHCl. This curve displays two zeros, corresponding to the temperatures of cold and heat denaturation. Also shown in the curve are the cooperative Gibbs free energies (ΔG^*) associated with the uncompensated exposure of apolar surfaces upon unfolding of each of the domains. For both domains, ΔG^* is positive for the heat denaturation and close to zero for the cold denaturation. This behavior results in a cooperative heat denaturation and a noncooperative cold denaturation.

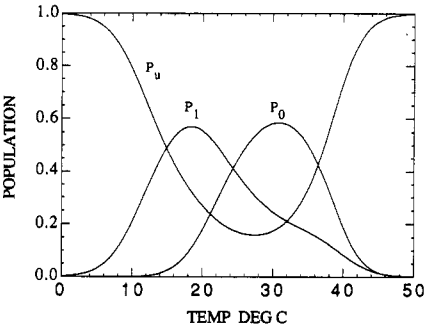


FIGURE 5: Temperature dependence of the population of molecules in the folded (P_0), unfolded (P_u), and partially folded intermediates (P_1 and P_2). P_1 is the intermediate in which the N domain is unfolded and the C domain folded. P_2 is the state in which the C domain is unfolded and the N domain is folded. The curves correspond to the experimental conditions in Figure 4.

to the uncompensated exposure of the N and C domains. As is the case with other proteins (Privalov & Gill, 1988), the overall free energy of stabilization displays a characteristic curvature with two temperatures (the cold and heat denaturation temperatures) at which it is equal to zero. Below and above these temperatures, the protein is in the denatured state.

As seen in the figure, the cooperative free energy is positive for the heat denaturation transition and close to zero for the cold denaturation transition. As a result, at the temperature of the heat denaturation transition, the Φ^* terms are equal to 0.08 and 0.09 for the N and C domains, respectively, whereas at the temperature of the cold denaturation transition they are equal to 0.99 and 0.98. These Φ^* values result in a small population of partially folded intermediates during the heat denaturation but a significant one during the cold denaturation. This is illustrated in Figure 5, where the population of states has been plotted as a function of temperature. This figure clearly shows that the total population of intermediates is less than 15% for the heat denaturation whereas it is close to 60% during the cold denaturation transition. It must be noted that the only intermediate that becomes populated is the one in which the N domain is unfolded and the C domain folded. The population of the other intermediate (N folded and C unfolded) is never greater than 10^{-6}. The population of unfolded molecules reaches 100% at low and high temperatures, consistent with the phenomenon of cold and heat denaturation. Under the conditions of these experiments (0.7 M GuHCl), the population of completely folded molecules never reaches 100% but exhibits a maximum near 30 °C where

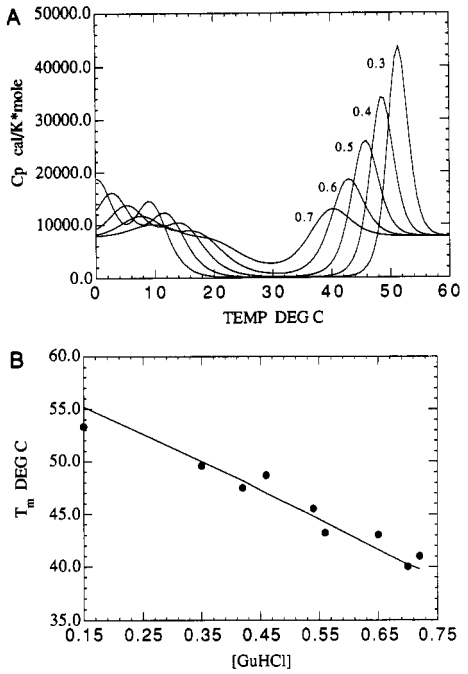


FIGURE 6: (A) Calculated dependence of the heat capacity function of phosphoglycerate kinase on the concentration of GuHCl. (B) Predicted and experimentally determined transition temperatures as a function of the concentration of GuHCl. All the experimental values were determined in DASM4 differential scanning calorimeters as described before (Griko et al., 1989; Galisteo et al., 1991).

it reaches about 60% of the total population.

The above calculations were performed for PGK at 0.7 M GuHCl, i.e., under conditions in which the folding/unfolding transition is thermodynamically reversible and both the cold and heat denaturation transitions are clearly visible in the calorimetric scans. As discussed before (Aune & Tanford, 1969; Ramsay & Freire, 1990), the dependence of the domain stability on [GuHCl] can be explicitly considered by linking the conformational stability equations with the denaturant binding equations. In the case of PGK, the N and C domain stability constants will assume the form:

$$K_i = K_i^0(1 + K_b a_G)^{\Delta n_i} \qquad (11)$$

where K_i^0 is the ith domain stability constant in the absence of denaturant, K_b the GuHCl binding constant, a_G the activity of GuHCl, and Δn_i the difference in the number of GuHCl binding sites between the unfolded and folded forms of the ith domain. Values for K_b have been previously estimated to be on the order of 0.6 M^{-1} for other proteins (Pace & Vanderburg, 1979; Ramsay & Freire, 1990). The binding enthalpy of GuHCl to proteins has been measured calorimetrically very recently and has been found to be about −2 kcal/mol (Makhatadze and Privalov, unpublished results). Using these binding parameters, Δn values of 45 and 43 GuHCl molecules are obtained for the N and C domains, respectively. These values are within the range expected for a protein of this size (Makhatadze and Privalov, unpublished results).

Figure 6A shows the predicted GuHCl concentration dependence of the excess heat capacity function of PGK. Figure 6B shows the theoretical and experimental dependence of the transition temperature on GuHCl concentration. The agreement with the experimental data obtained under conditions of thermodynamic reversibility is excellent. At zero GuHCl concentration and pH 6.5, our calculations predict a T_m of 59 °C, which is about 4 °C higher than the experimental values obtained in the pH range 6–7 with protein concentrations of about 1 mg/mL and higher (Hu & Sturtevant, 1987; Brandts et al., 1989; Galisteo et al., 1991). It must be noted, however, that the thermal denaturation of PGK at zero GuHCl concentration is irreversible and strongly rate-limited. In addition, the rate of irreversible denaturation of PGK depends on protein concentration, and as a result, the T_m values of the calorimetric transitions are higher at protein concentrations below 1 mg/mL. It is interesting to note that the T_m value obtained at 0.12 mg/mL (where the rate of irreversible denaturation is slow) is about 59–60 °C, in good agreement with the predicted value. The predicted overall enthalpy change under these conditions is 217 kcal/mol, in excellent agreement with the values of 223 kcal/mol reported by Hu and Sturtevant (1987) and 204 kcal/mol reported by Galisteo et al. (1991).

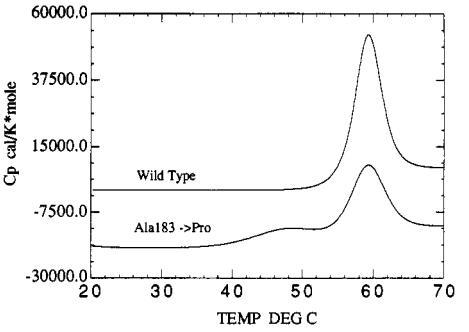


FIGURE 7: Calculated heat capacity function for the wild type and Ala183→Pro mutant of yeast phosphoglycerate kinase. As shown experimentally by Bailey et al. (1990), the transition of this mutant exhibits two peaks centered around 44 and 54 °C corresponding to the N and C domains, respectively. The curves have been displaced in the vertical axis for display purposes.

The calculations presented in this paper indicate that the N domain is considerably less stable than the C domain and that it is effectively stabilized by its interactions with the C domain. In fact, the only partially folded state that ever becomes significantly populated is the one in which the N domain is unfolded and the C domain folded (see Figure 5). These results are consistent with folding kinetics data (Betton et al., 1985), with partial proteolysis experiments (Betton et al., 1989), and with ligand perturbation experiments (Brandts et al., 1989) indicating that the C domain is the most stable and able to fold by itself. Also, according to the crystallographic structure (Watson et al., 1982), there are 102 hydrogen bonds in the N domain compared to 150 in the C domain. These numbers yield an average of 0.54 hydrogen bond per amino acid for the N domain versus 0.66 for the C domain. The average number for small globular proteins is 0.73 (Privalov & Gill, 1988). The higher stability of the C domain is primarily a consequence of the larger number of hydrogen bonds per residue, since the normalized exposure of hydrophobic surfaces upon unfolding is similar for both domains.

Finally, Bailey et al. (1990) have recently produced a mutant of yeast PGK in which alanine-183 has been replaced by a proline. Alanine-183 is located in the hinge region between the two domains. It has been argued that this region of the molecule might be important for hinge-bending domain movement (Watson et al., 1982). This mutant form of the enzyme exhibits a reduced cooperative behavior characterized by the presence of two peaks in the heat denaturation transition. The lower temperature peak corresponds to the N domain and the high-temperature peak to the C domain. A minimal perturbation mechanism consistent with this behavior involves the exposure to the solvent of the hydrophobic interface and the disruption of four hydrogen bonds in the hinge region of the protein. Figure 7 shows the heat capacity profile

predicted by the imposition of those conditions. As seen in the figure, for the mutant protein the main transition peak is splitted into two peaks: a low-temperature peak centered at 45 °C, corresponding to the unfolding of the N domain, and the main peak, still centered at 59 °C and corresponding to the C domain. This theoretically predicted curve is in excellent agreement with the experimentally observed calorimetric profile, also characterized by the appearance of a second, low-temperature peak corresponding to the N domain and located 10 °C below the main transition peak (Bailey et al., 1990).

The calculations presented in this paper as well as those presented earlier for myoglobin (Freire & Murphy, 1991) constitute the first attempts to predict the calorimetrically observed folding/unfolding behavior of proteins from structural information. The development of the ability to predict the stability and cooperative folding behavior of proteins from crystallographic data should improve existing strategies for the rational design and modification of proteins.

Registry No. PGK, 9001-83-6; GuHCl, 50-01-1.

REFERENCES

Abola, E. E., Bernstein, F. C., Bryant, S. H., Koetzle, T. F., & Weng, J. (1987) in *Crystallographic Databases—Information Content, Software Systems, Scientific Applications* (Allen, F. H., Bergerhoff, G., & Sievers, R., Eds.) p 107, Data Commission of the International Union of Crystallography, Bonn/Cambridge/Chester.

Aune, & Tanford, C. (1969) *Biochemistry 8*, 4586–4590.

Bailey, J. M., Lin, L. N., Brandts, J. F., & Mas, M. T. (1990) *J. Protein Chem 9*, 59–67.

Baldwin, R. L. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 8069–8072.

Banks, R. D., Blake, C. C. F., Evans, P. R., Haser, R., Rice, D. W., Hardy, G. W., Merrett, M., & Philips, R. W. (1979) *Nature 279*, 773–777.

Betton, J. M., Desmadril, M., Mitraki, A., & Yon, J. M. (1985) *Biochemistry 24*, 4570–4577.

Betton, J. M., Desmadril, M., & Yon, J. M. (1989) *Biochemistry 28*, 5421–5428.

Brandts, J. F., Hu, C. Q., Lin, N. L., & Mas, M. T. (1989) *Biochemistry 28*, 8588–8596.

Chothia, C. (1976) *J. Mol. Biol. 105*, 1–14.

Connolly, M. L. (1983a) *Science 221*, 709–713.

Connolly, M. L. (1983b) *J. Appl. Crystallogr. 16*, 548–558.

Freire, E. (1990) *Comments Mol. Cell. Biophys. 6*, 123–140.

Freire, E. & Biltonen, R. L. (1978) *Biopolymers 17*, 463–479.

Freire, E., & Murphy, K. P. (1991) *J. Mol. Biol.* (in press).

Galisteo, M. L., Mateo, P. L., & Sanchez-Ruiz, J. M. (1991) *Biochemistry 30*, 2061–2066.

Goins, B., & Freire, E. (1988) *Biochemistry 27*, 2046–2052.

Griko, Y. V., Venyaminov, S. Y., & Privalov, P. L. (1989) *FEBS Lett. 244*, 276–278.

Hu, C. Q., & Sturtevant, J. M. (1987) *Biochemistry 26*, 178–182.

Lee, B., & Richards, F. M. (1971) *J. Mol. Biol. 55*, 379–400.

Murphy, K. P. (1990) Ph.D. Thesis, The University of Colorado.

Murphy, K. P., & Gill, S. J. (1990) *Thermochim. Acta 172*, 11–20.

Murphy, K. P., & Gill, S. J. (1991) *J. Mol. Biol.* (in press).

Murphy, K. P., Privalov, P. L., & Gill, S. J. (1990) *Science 247*, 559–561.

Pace, C. N., & Vanderburg, K. E. (1979) *Biochemistry 18*, 288–292.

Pfeil, W., & Privalov, P. L. (1976) *Biophys. Chem. 4*, 33–40.

Privalov, P. L. (1979) *Adv. Protein Chem. 33*, 167–241.

Privalov, P. L. (1982) *Adv. Protein Chem. 35*, 1–104.

Privalov, P. L., & Khechinashvilli, N. N. (1974) *J. Mol. Biol. 86*, 665–684.

Privalov, P. L., & Gill, S. J. (1988) *Adv. Protein Chem. 39*, 191–234.

Privalov, P. L., Griko, Y. V., Venyaminov, S. Y., & Kutyshenko, V. P. (1986) *J. Mol. Biol. 190*, 487–498.

Ramsay, G., & Freire, E. (1990) *Biochemistry 29*, 8677–8683.

Ramsay, G., Montgomery, D., Berger, D., & Freire, E. (1989) *Biochemistry 28*, 529–533.

Scholtz, J. M., Marqusee, S., Baldwin, R. L., York, E. J., Stewart, J. M., Santoro, M., & Bolen, D. W. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 2854–2858.

Watson, H. C., Walker, N. P. C., Shaw, P. J., Bryant, P. L., Fothergill, L. A., Perkins, R. E., Conroy, S. C., Dobson, M. J., Tuite, M. F., Kingsman, A. J., & Kingsman, S. M. (1982) *EBMO J. 1*, 1635–1640.