Thermodynamic Properties of Transient Intermediates and Transition States in the Folding of Two Contrasting Protein Structures†

Martin J. Parker,* Mark Lorch, Richard B. Sessions, and Anthony R. Clarke

Department of Biochemistry, University of Bristol, School of Medical Sciences, University Walk, Bristol BS8 1TD, U.K.

Received October 3, 1997; Revised Manuscript Received December 4, 1997

ABSTRACT: The N-terminal domain of phosphoglycerate kinase (N-PGK) and domain 1 of the T-cell adhesion protein CD2 (CD2.d1) fold through rapidly formed and transiently populated intermediate states in reactions which have no kinetic complications arising from proline isomerization or disulfide bonding. We have evaluated the thermodynamic parameters (ΔC_p, change in heat capacity; ΔS, entropy change; ΔH, enthalpy change) for each experimentally accessible step in these folding reactions. Despite their different topologies and amino acid compositions, the individual steps [U−I (unfolded to intermediate state), I−t (intermediate to major transition state), and t−F (transition state to the fully folded state)] have closely similar qualitative properties in the two proteins. For both, the heat capacity changes are proportional to *m*-value changes (Δm) for every step in the reaction, but the ratio $\Delta C_p/\Delta m$ is lower for N-PGK, presumably owing to a much larger compliment of aromatic amino acids in the core. According to measurements of ΔC_p and Δm, the I-states are highly condensed (65−70% for N-PGK and 40−45% dehydrated for CD2.d1), yet the changes in entropy in the U-to-I transition are small, showing that the entropy gained from desolvation must be balanced by that lost in ordering the chain. The high degree of conformational order in the I-state, implied by these measurements, is mirrored by the extensive, native secondary structure revealed by amide exchange measurements [Hosszu, L. L. P., et al. (1997) *Nat. Struct. Biol. 4*, 801−804; Parker, M. J., et al. (1997) *Biochemistry 36*, 13396−13405]. At 25 °C the transition state barrier has an entirely enthalpic origin, the entropic contribution being favorable. The latter observation implies that, during the consolidation of structure occurring in the I-to-F step, further dehydration (positive ΔS) precedes side-chain locking (negative ΔS). Only after the transition state is surmounted do we see a net entropic penalty arising from the widespread ordering of side chains.

There is a distinct lack of consensus in the way we view the properties of states which arise in protein folding reactions, and it is unlikely that a single set of experiments performed on a single protein will clarify our picture of the process. Rather, it is necessary to amass a body of experimental information so that the common underlying features can be identified and accounted for. In general there are four experimentally accessible states in folding reactions: the fully unfolded state (U), a rapidly formed intermediate (I), the rate-limiting transition state (t), and the folded state (F). In some cases, factors such as domain docking, subunit association, proline isomerization, or the breaking and making of disulfide bonds contribute to the folding dynamics; and in small proteins the I-state is sometimes not seen, due either to its absence or to its lack of population (*1*, *2*).

To understand the initial events in the pathway, the structural, energetic, and dynamic properties of the I- and t-states must be elucidated. Our structural knowledge of the I-state has been drawn largely from circular dichroism, which measures the quantity of ordered secondary structure (*3*), and from pulse-protection, which measures resistance to exchange of amide hydrogen atoms, which in turn is taken to be a reporter of hydrogen bond formation (*4*, *5*). These methods reveal that a large degree of the native state secondary structure has been established at this early stage in the process, before the major kinetic barrier is surmounted. Such results have been reinforced by experiments which measure the effect of mutations on the stability of I-states. These reveal that contacts which develop in the transition state are already present, in a weakened form, in the rapidly formed intermediate (*6*−*8*).

Thus far, the most effective way of elucidating the properties of the rate-limiting t-state is by a combination of site-directed mutagenesis and rate measurements which report the degree of inter-residue contact in the transition state (*7*−*10*). These studies show that many of the core, hydrophobic contacts are partially formed, or fully formed in only a proportion of those molecules which constitute the transition state population, i.e., no side chain is in a native environment in all members of the ensemble at this stage of the reaction.

A further level of probing the system experimentally is to derive classical thermodynamic characteristics by examining how the kinetics of folding reactions are influenced by temperature. Because the system of classical thermodynamics is developed from studies of heat engines, it is notoriously difficult to interpret the parameters of enthalpy, entropy, and heat capacity on the scale of a complex polymer folding in

† This work was supported by a project grant from the BBSRC (U.K.), and equipment funding was from the Wellcome Trust. A.R.C. is a Lister Institute research fellow.

* Author to whom correspondence should be addressed.

S0006-2960(97)02459-8 CCC: $15.00
Published on Web 01/30/1998

an aqueous environment. Nonetheless, largely owing to an empirical database of the behavior of small molecules (*11*−*14*, and references therein), it is known that the removal of nonpolar groups from contact with the solvent is accompanied by a reduction in heat capacity and that there are two large and opposite contributions from entropy. A massive reduction in conformational entropy of the main chain and side chains during folding is balanced by the high entropy of dehydration when side chains become excluded from contact with the solvent. The enthalpic component is determined by the balance of polar interactions (largely hydrogen bonding), van der Waals attractions, and steric repulsions.

Changes in ΔC_p, ΔS, and ΔH have been measured for the folding of a few proteins, and in these cases emphasis has been placed on the properties of the transition state (*15*−*20*). In this paper we add to this body of experimental work by examining the influence of temperature on the folding kinetics of N-PGK,[1] a large (175-residue) $\alpha\beta$ protein (*21*), and CD2.d1, a 98-residue all-β protein (*22*). In particular, we examine the behavior of the rapidly formed intermediate, which accumulates before the rate-limiting barrier is traversed (*23*, *24*), and the nature of the rate-limiting barrier between the intermediate and the native state. Both proteins are single-domain structures, and their folding kinetics are uncomplicated by proline isomerizations or disulfide bond rearrangements. In addition, the structural properties of the I-states of both proteins have been characterized by amide exchange studies (*25*, *26*).

EXPERIMENTAL PROCEDURES

Source of Protein. The N-terminal domain of PGK from *Bacillus stearothermophilus* (N-PGK; residues 1−175) was expressed and purified as described previously (*23*). Domain 1 of rat CD2 (CD2.d1; residues 1−98) was expressed and purified as described previously (*24*).

Protein concentrations were estimated by UV absorption of aromatic residues at 280 nm (ϵ = 5500 M^{-1} cm^{-1} for tryptophan and 1100 M^{-1} cm^{-1} for tyrosine).

Folding and Unfolding Rates. For the folding rates of N-PGK, a 20 μM solution of unfolded protein, containing 50 mM triethanolamine (TEA; pH 7.5), 2 M guanidine hydrochloride (GuHCl), and 2 mM dithiothreitol (DTT), was mixed against 10 volumes of buffer or buffered GuHCl solution, containing 2 mM DTT, in a Hi-Tech SF-51 stopped-flow apparatus at the appropriate temperature. An excitation wavelength of 270 nm was selected by a single monochromator (slit width 5 nm) from a mercury−xenon light source, and the total fluorescence emission intensity was recorded as a function of time. For the unfolding rates, a 20 μM solution of N-PGK, in the above buffer with the exclusion of GuHCl, was mixed against an equal volume of a buffered GuHCl solution containing 2 mM DTT at the appropriate temperature, using the same apparatus and optical setup.

For the folding rates of CD2.d1, a 10 μM solution of unfolded protein, containing 50 mM sodium acetate (pH 4.5) and 2 M GuHCl, was mixed against 10 volumes of buffer or buffered GuHCl at the appropriate temperature in an Applied Photophysics stopped-flow apparatus. An excitation wavelength of 288 nm was selected by a single monochromator (slit width, 5 nm) from a mercury−xenon light source, and the fluorescence intensity above 320 nm was recorded using an emission cutoff filter. For the unfolding reactions, a 10 μM solution of CD2.d1, in the above buffer, was mixed against 10 volumes of buffered GuHCl solution at the appropriate temperature, and the reaction was recorded as above.

Owing to technical limitations of the stopped-flow apparatus (loss of optical signals at high temperatures and imperfect sealing at low temperatures), accurate rate constants could only be obtained for N-PGK between 15 and 35 °C and for CD2.d1 between 15 and 45 °C.

Solubility Measurements. Saturated solutions of *N*-acetyltryptophanamide (NATA) (Sigma) were made up in water, containing an appropriate concentration of GuHCl. The solutions were sonicated and then incubated for 48 h at the appropriate temperature with vigorous shaking, to allow complete equilibration. Insoluble material was then removed by centrifugation. The concentration of soluble NATA in each sample was then assessed by measuring the absorbance at 280 nm at the appropriate temperature.

All reaction solutions were maintained at the appropriate temperature using thermostated circulating water baths and were monitored continuously with a sensitive thermocouple. From this, the fluctuation in temperature was determined to be no more than ±0.1 °C.

ANALYTICAL PROCEDURES

Treatment of Kinetic Data. The transients of fluorescence intensity (I) versus time, which are single relaxation processes in both the folding and unfolding processes of N-PGK and CD2.d1, were fitted to the equation $I = I_a[1 - \exp(-kt)] + I_o$ for rising intensities (where I_a is the fluorescence amplitude of the reaction, k is the observed rate constant for the relaxation, and I_o is the initial intensity) and $I = I_a \exp(-kt) + I_f$ for decreasing intensities (where I_f is the final fluorescence intensity).

The rate profiles [observed rate constant (k_{obs}) vs denaturant activity (D; see below)] were fitted to the following equation (23):

$$k_{obs} = k_{F-I(w)} \exp(-m_t D) + \frac{k_{I-F(w)} \exp((m_I - m_t)D)}{[1 + 1/(K_{I/U(w)} \exp((m_U - m_I)D))]} \quad (1)$$

where $k_{I-F(w)}$ and $k_{F-I(w)}$ are the folding and unfolding rates in water, respectively, associated with the rate-limiting intermediate state (I) to folded state (F) transition, and $K_{I/U(w)}$ is the equilibrium constant in water for the rapid unfolded (U) to intermediate transition. The m-values describe how the free energies of the states (denoted by the subscripts; where t is the transition state associated with the I to F transition) vary as a function of D.

The free energy changes associated with the equilibrium transitions were calculated as follows: U to F, $\Delta G_{F-U} = -RT \ln(k_{I-F(w)} K_{I/U(w)}/k_{F-I(w)})$; I to F, $\Delta G_{F-I} = -RT \ln(k_{I-F(w)}/k_{F-I(w)})$; I to U, $\Delta G_{I-U} = -RT \ln K_{I/U(w)}$, where T is the temperature and R is the gas constant.

[1] Abbreviations: CD2.d1, domain 1 of the cell surface receptor protein CD2 from rat; DTT, dithiothreitol; GuHCl, guanidine hydrochloride; NATA, *N*-acetyltryptophanamide; N-PGK, N-terminal domain of phosphoglycerate kinase from *Bacillus stearothermophilus*; TEA, triethanolamine hydrochloride.

Variation of Denaturant Activity with Temperature. For the analysis of the kinetic data, GuHCl concentration ([GuHCl]) is converted to denaturant activity (D) according to the following equation:

$$D = K_{den}[\text{GuHCl}]/(K_{den} + [\text{GuHCl}]) \quad (2)$$

where K_{den} is a denaturation constant with the value 7.5 M at 25 °C (*23*). To examine the influence of temperature on the molar ability of GuHCl to solvate hydrocarbon (the variation of K_{den} with temperature), we measured the solubility of NATA as a function of [GuHCl] at different temperatures (see above). For each temperature, the difference in the free energy of solvation (ΔG_s) for NATA between water and a given concentration of GuHCl is given by

$$\Delta G_s = -RT \ln(A_{280}/A_{280(w)}) \quad (3)$$

where A_{280} is the absorbance at 280 nm measured in a particular concentration of GuHCl and $A_{280(w)}$ is that measured in water. Plots of ΔG_s vs [GuHCl] were fitted to the hyperbolic relationship

$$\Delta G_s = \Delta G_{s,max}[\text{GuHCl}]/(C_{0.5} + [\text{GuHCl}]) \quad (4)$$

where $\Delta G_{s,max}$ is the notional maximum change in the free energy of solvation at an infinite concentration of denaturant and $C_{0.5}$ is a denaturation constant with units of M.

Variation of Free Energy with Temperature. For the variation of the free energy change associated with a particular transition (ΔG) with temperature (T), data were fitted to the equation

$$\Delta G_{(T)} = \Delta H_{(T_o)} + \Delta C_p(T - T_o) - T\Delta S_{(T_o)} - \Delta C_p T \ln(T/T_o) \quad (5)$$

where $\Delta H_{(T_o)}$ and $\Delta S_{(T_o)}$ are the enthalpy and entropy changes, respectively, at an arbitrarily defined reference temperature (T_o) and ΔC_p is the change in heat capacity (at constant pressure) associated with the transition.

Variation of the Unfolding Rate with Fractional Solvent Exposure. The natural logarithm of the rate-limiting unfolding rate constant ($\ln k_u$) was plotted against the fractional solvent exposure of the transition state ($\alpha^{\ddagger} = m_t/m_U$) for the following proteins: C-terminal domain of PGK and six core hydrophobic side chain deletion (CHSD) mutants of this protein (*8*), N-PGK (*23*), the α and β domains of hen lysozyme (*23*), CD2.d1 and CHSD mutants thereof (M. Lorch, M. J. Parker, and A. R. Clarke, unpublished data), CI2 and CHSD mutants thereof (*27*, *28*), barnase and CHSD mutants thereof (*29*, *30*), ubiquitin and CHSD mutants thereof (*7*), cheY (*31*), cspB (*18*), RNase HI (*32*), and P22 arc repressor and CHSD mutants thereof (*33*).

All data were fitted using the GraFit analysis software (Erithracus Software, U.K.).

RESULTS AND DISCUSSION

Treatment of Data. The denaturant dependencies of the observed relaxation rates for folding/unfolding of N-PGK and CD2.d1 at different temperatures are shown in panels a and b of Figure 1, respectively. As described previously (*23*, *24*), to correct for the nonlinear dependence of the free energy of protein folding on the concentration of GuHCl, the rate constants are plotted against denaturant activity (D), calculated using eq 2. Where high denaturant concentrations are required to unfold a protein, this linearized scale allows a more reliable extrapolation of data to a condition where denaturant is absent. This is particularly important for the evaluation of unfolding rates and equilibrium constants in water and for determining proper m-values. Reassuringly, a direct calorimetric evaluation of the dependence of the free energy change of protein unfolding on denaturant concentration (*34*) provides a molar denaturation scale in very close agreement with that derived from solubility data (*23*). In addition, the use of a denaturant activity scale provides a value for the rate-limiting unfolding rate of CD2.d1 ($k_{F-I(w)}$) that is in very close agreement with the observed rate of exchange for amide protons in native conditions at the EX1 limit, studied by NMR (M. J. Parker, and A. R. Clarke, unpublished data). Here, for amide protons which are fully protected in F, exchange is limited by the global opening rate. If raw concentration is used, these rates differ by more than one order of magnitude.

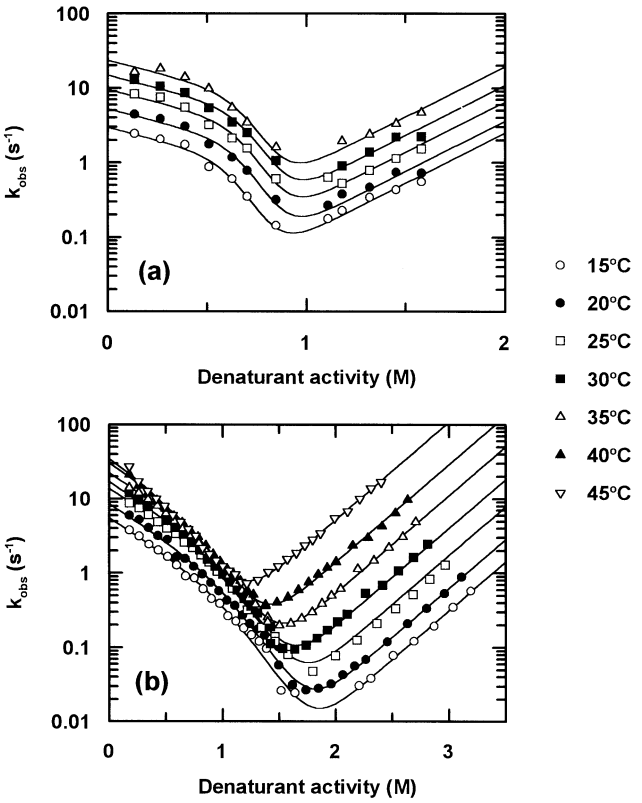

FIGURE 1: Folding dynamics of N-PGK and CD2.d1. Shown in (a) and (b) are the observed relaxation rates for folding/unfolding (k_{obs}) of N-PGK and CD2.d1, respectively, determined by stopped-flow fluorimetry at different temperatures (see Experimental Procedures). Here, k_{obs} is plotted against denaturant activity (D) (see Analytical Procedures and legend to Figure 2). The continuous curves represent fits to eq 1 (Analytical Procedures).

At 25 °C a value of 7.5 M is used for the denaturation constant (K_{den}; *23*). To examine the possibility that temperature may alter the ability of GuHCl to solvate protein hydrocarbon moieties (and thus alter the value of K_{den}), we have measured the GuHCl dependence of the free energy change of solvation of *N*-acetyltryptophanamide (NATA) over the temperature range used in the kinetic measurements (Figure 2a). This compound was chosen because its $C_{0.5}$

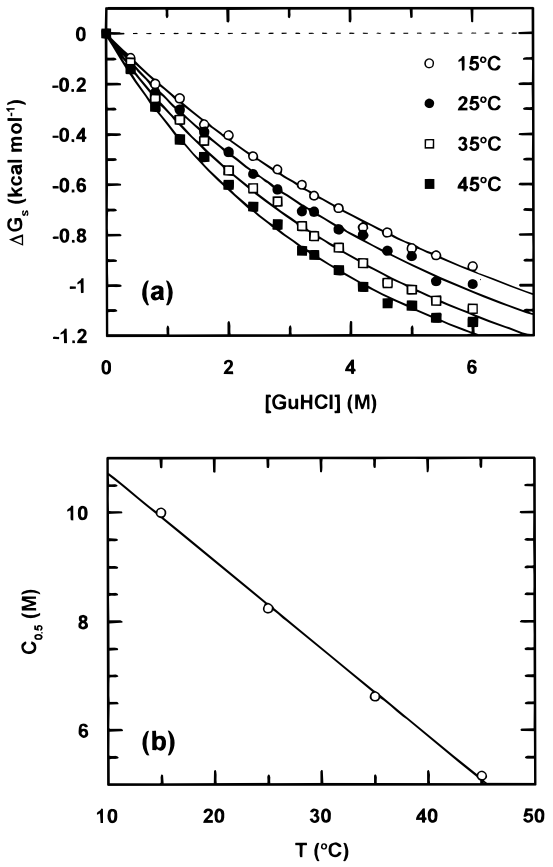

FIGURE 2: Denaturant activity versus temperature. In (a) the free energy change of solvation (ΔG_s) of NATA from water to increasing concentrations of GuHCl has been plotted at 15, 25, 35, and 45 °C (see Experimental Procedures). The data have been fitted to eq 4 (Analytical Procedures) (fits shown as continuous curves). The calculated values for $C_{0.5}$ at 15, 25, 35, and 45 °C are 10.0, 8.24, 6.62, and 5.16 M, respectively (average error, 20%). The calculated values for $\Delta G_{s,max}$ are −2.52, −2.43, −2.35, and −2.22 kcal mol^{-1}, respectively (average error, 20%). From 15 to 45 °C, $C_{0.5}$ varies linearly with temperature. This is illustrated by the plot in (b). The straight line fit ($C_{0.5(T)} = C_{0.5(0)} + mT$) yields the following parameters: $C_{0.5(0)} = 12.3 \pm 0.2$ M and $m = -0.16 \pm 0.05$ M per °C. For the protein denaturant activity scale we use a value for K_{den} of 7.5 M (see Analytical Procedures and ref *23*). To calculate K_{den} at other temperatures we use the following relationship: $K_{den(T)} = (7.5/C_{0.5(25)})\ C_{0.5(T)}$.

value (a measure of the degree of nonlinearity between concentration and activity) is close to the average for all amino acid residues (*23*); thus its behavior is representative, and its concentration can be accurately measured by absorbance. The data have been fitted to the empirical relationship given by eq 4 to yield values of $\Delta G_{s,max}$ (the maximum change in solvation free energy at an infinite concentration of denaturant) and $C_{0.5}$ (the concentration of denaturant required to achieve half this maximal change in solvation free energy). The calculated values of $\Delta G_{s,max}$ and $C_{0.5}$ for NATA at 15, 25, 35, and 45 °C are given in the legend to Figure 2. Over the experimental temperature range used in the kinetic measurements, the value of $C_{0.5}$ varies linearly with temperature (Figure 2b). The parameters describing the straight-line fit in Figure 2b have been used to calculate the denaturant activity scales for the kinetic data in Figure 1, as described in the legend to Figure 2.

If the temperature variation of $C_{0.5}$ is not taken into account the *m*-values associated with the F-to-t (m_t), F-to-I (m_I), and F-to-U (m_U) transitions vary linearly with temperature. The linear responses of m_t, m_I, and m_U to temperature are identical (data not shown). As these *m*-values provide a qualitative measure of the difference in the solvation of hydrocarbon (*23*, *24*, *35*−*37*), one would conclude that, as the temperature changes, the degree of hydrocarbon exposure of the t, I, and U states, relative to the folded state, changes by exactly the same amount, irrespective of their structural and energetic properties. This seems unlikely. A more probable explanation is that the molar ability of the denaturant to alter the free energy difference between the F, t, I, and U states (change the free energy of solvation of hydrocarbon) changes with temperature. This conclusion is evidenced by the data in Figure 2. Moreover, when this temperature variation in $C_{0.5}$ is accounted for, the *m*-values are the same, within error, at each temperature.

It is important to establish, however, how the error in $C_{0.5}$ affects the extrapolated rate and equilibrium values, and consequently the derived temperature-dependent parameters. The error in $C_{0.5}$ is expected to affect, most notably, the value of $k_{F-I(w)}$ for CD2.d1, as evaluation of this parameter requires the longest extrapolation. Consequently, the enthalpy (ΔH), entropy (ΔS), and heat capacity (ΔC_p) changes associated with the F-to-t transition of CD2.d1 are likely to be affected most by the error in the calculated $C_{0.5}$ values. Using upper and lower limiting values for $C_{0.5}$ at each temperature, obtained from the 95% confidence limits of the fits in Figure 2, the absolute error in $k_{F-I(w)}$ is estimated to be no more than 20%. This uncertainty produces estimates for the absolute errors in ΔH, ΔS, and ΔC_p, incurred by the uncertainty in $C_{0.5}$, of 2, 5, and 14%, respectively. These errors are relatively small and do not alter the qualitative conclusions presented here.

Transitions between Ground States. The kinetic data in Figure 1 have been fitted to eq 1, which describes a three-state kinetic mechanism (i.e., U to I to F), to yield values for $k_{I-F(w)}$, $k_{F-I(w)}$, $K_{I/U(w)}$, m_U, m_I, and m_t (see Analytical Procedures). The temperature dependencies of the free energy changes associated with the equilibrium transitions (ΔG_{F-U}, ΔG_{F-I}, and ΔG_{I-U}; see Analytical Procedures) for N-PGK and CD2.d1 are plotted in panels a and b of Figure 3, respectively. The data have been fitted to eq 5 to yield values of ΔH, ΔS, and ΔC_p for each ground state in the folding reactions of N-PGK and CD2.d1. These values are shown in Table 1.

While at first sight it appears odd to use kinetic rather than equilibrium data to examine transitions between ground states, the measurement of the temperature dependence of folding and unfolding rates has distinct methodological advantages. First, it overcomes the problem of aggregation of non-native states frequently encountered in scanning calorimetry, but more importantly it offers the possibility of measuring thermodynamic parameters for the U-to-I transition which cannot normally be explored by equilibrium unfolding methods as the intermediate is a transient species.

The net entropic and enthalpic changes associated with the U-to-I and I-to-F transitions are the result of combining massive favorable and unfavorable energetic contributions which almost cancel. This is well illustrated by crude estimates of the entropic factors in a folding reaction. The desolvation entropy for removal of water from a core side chain ($-T\Delta S$) at 25 °C will contribute around −3 kcal mol^{-1}

Table 1: Thermodynamic Parameters for the Folding Reactions of N-PGK and CD2.d1 at 25 °C

	U−I	I−F	I−t	t−F
		N-PGK (T_o = 25 °C)		
ΔC_p (kcal mol^{-1} K^{-1})	−1.30 (−1.10, −1.50)	−0.73 (−0.67, −0.79)	−0.28 (−0.20, −0.36)	−0.45 (−0.39, −0.51)
Δm[a] (M^{-1})	−11.1 (−8.36, −13.8)	−4.7 (−4.1, −5.3)	−1.6 (−1.2, −2.0)	−3.1 (−2.5, −3.7)
ΔH (kcal mol^{-1})	4.62 (5.17, 4.07)	−0.20 (−0.01, −0.40)	18.2 (18.4, 18.0)	−18.4 (−18.2, −18.6)
$-T\Delta S$ (kcal mol^{-1})	−9.24 (−8.65, −9.83)	−3.78 (−3.60, −3.96)	−5.96 (−5.72, −6.20)[b] −10.1 (−9.86, −10.3)[c]	2.18 (2.36, 2.00)[b] 6.26 (6.44, 6.08)[c]
		CD2.d1 (T_o = 25 °C)		
ΔC_p (kcal mol^{-1} K^{-1})	−0.86 (−0.72, −1.00)	−1.04 (−0.84, −1.24)	−0.26 (−0.10, −0.42)	−0.78 (−0.51, −1.05)
Δm[a] (M^{-1})	−3.5 (−2.9, −4.5)	−5.2 (−4.8, −5.6)	−2.2 (−1.8, −2.6)	−3.0 (−2.6, −3.4)
ΔH (kcal mol^{-1})	−4.67 (−3.83, −5.51)	−18.8 (−17.4, −20.2)	12.6 (13.6, 11.6)	−31.4 (−29.6, −33.2)
$-T\Delta S$ (kcal mol^{-1})	2.50 (3.38, 1.62)	12.2 (13.2, 11.2)	−0.57 (0.50, −1.63)[b] −4.77 (−3.71, −5.83)[c]	12.8 (13.0, 12.6)[b] 6.9 (17.1, 16.7)[c]

[a] Δm values are averaged from the rate profiles at different temperatures. [b] A value for $k'_{(T)}$ of 1×10^{10} s^{-1} used to calculate $\Delta S^{\ddagger}$. [c] A value for $k'_{(T)}$ of 1×10^{7} s^{-1} used to calculate $\Delta S^{\ddagger}$ (see Results and Discussion). Upper and lower limits for these parameters, based on a confidence limit of 95%, are given in parentheses and were calculated using standard statistical formulas and tables. The uncertainty in the calculated $C_{0.5}$ values is estimated to affect the ΔH, ΔS, and ΔC_p values by no more than 2, 5, and 14%, respectively (see Results and Discussion).

to the free energy of folding, with a further contribution from the associated peptide unit of around −4 kcal mol^{-1} (*13*). Hence, for a 100-amino acid protein with a core comprising, say, 40 residues the internal entropy change due to desolvation will be around −280 kcal mol^{-1}. However, the configurational entropy loss for the backbone alone upon folding costs around 150 kcal mol^{-1} [+1.5 kcal mol^{-1} per residue at 25 °C (*38*)]. With a similar entropic cost for reducing the configurational freedom of side chains, it can be appreciated that the net entropy changes in the folding reactions (−13 kcal mol^{-1} for N-PGK and +15 kcal mol^{-1} for CD2.d1) are small by comparison to the magnitude of these opposing contributions. A similar argument is applicable to enthalpy, with very large positive ΔH values for desolvation of internal residues being matched by the favorable enthalpic contribution from backbone hydrogen bonding and from van der Waals and solvent−solvent interactions in the folded state.

Although molecular interpretations of classical thermodynamic parameters must be approached with some trepidation, the changes in heat capacity, entropy, and enthalpy give some indication of the underlying processes in the reaction. Of these parameters, the most simply interpreted from an empirical viewpoint is the change in heat capacity, ΔC_p, which has been shown from studies of model compounds to equate with the exclusion of nonpolar groups from the solvent (*39*−*43*). As the *m*-value is argued to reflect the same phenomenon (*35*−*37*, *44*, *45*), then ΔC_p and Δm should be highly correlated, an expectation which is confirmed by the plot in Figure 4. These parameters show that, for both proteins, a large proportion of the overall exclusion of hydrophobic groups from the bulk solvent which occurs during folding is achieved upon formation of the I-state (65−70% for N-PGK and 40−45% for CD2.d1).

It is interesting to note that the slope of the plot of Δm vs ΔC_p is smaller for N-PGK than for CD2.d1 and that, while Δm is simply proportional to the amount of hydrocarbon becoming exposed in a transition (*23*, *24*), ΔC_p is different for the solvation of aromatic and aliphatic groups (*11*). Hence, one contributory factor is that the N-PGK core has a greater aromatic content. However, the difference in slope is too large to be accounted for by this effect alone, and it also may be true that the unfolding of N-PGK results in the solvation of relatively more polar groups than CD2.d1 [ΔC_p values for the solvation of polar groups are positive (*11*)].

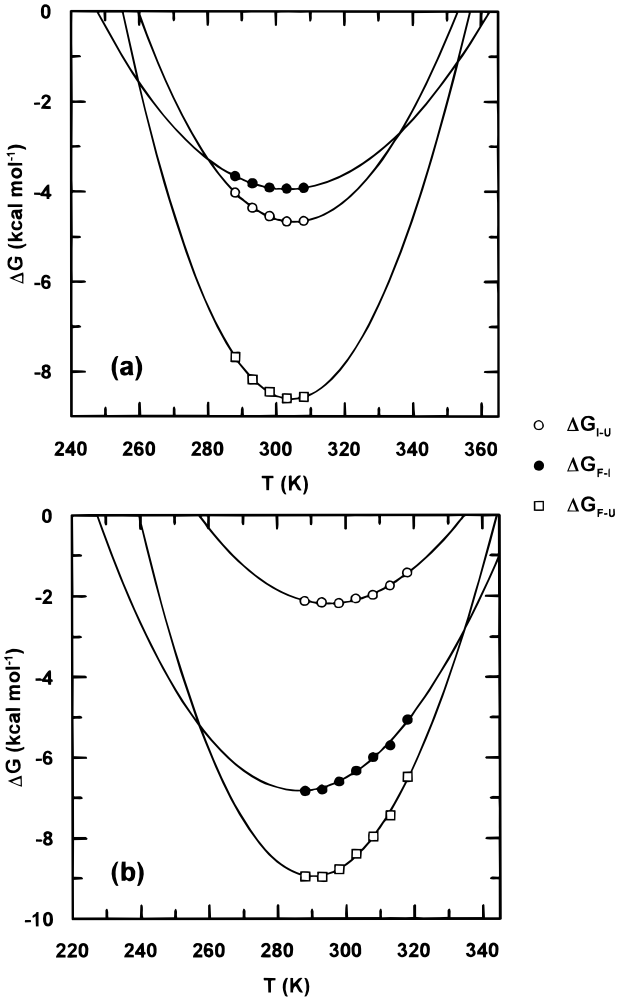

FIGURE 3: Variation of ΔG with temperature: equilibrium transitions. The free energy changes associated with the U-to-I (ΔG_{I-U}), I-to-F (ΔG_{F-I}), and U-to-F (ΔG_{F-U}) equilibrium transitions (see Analytical Procedures) are plotted as a function of temperature for N-PGK and CD2.d1 in (a) and (b), respectively. The continuous curves represent fits to eq 5 (Analytical Procedures). The calculated values for the enthalpy, entropy, and heat capacity changes associated with these transitions are given in Table 1.

As might be expected for a protein from a thermophilic organism, N-PGK attains its maximum stability (i.e., the minimum value of ΔG_{F-U}) at a higher temperature than does CD2.d1 (31 °C compared to 17 °C). The plots of ΔG vs T

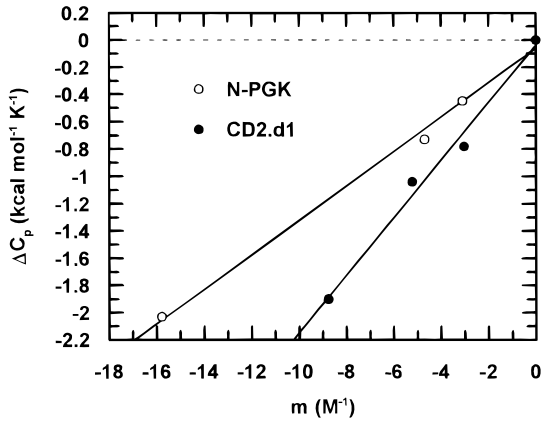

FIGURE 4: Correlation between Δm and ΔC_p. The heat capacity changes for the U, I, and t states, relative to F, are plotted here against their respective *m*-values (also relative to F), for N-PGK and CD2.d1. Both sets of data are highly correlated (correlation coefficients for both are greater than 0.98). The straight line fits yield slopes of 0.13 (N-PGK) and 0.21 (CD2.d1) kcal mol^{-1} K^{-1} M.

in Figure 3 also show that there is no temperature at which the I-state is the predominant species, thus precluding its characterization by scanning calorimetry.

A clear observation we make from the properties of the I-state is that it has undergone a large reduction in heat capacity when compared to the unfolded state. This demonstrates that a large degree of desolvation has occurred in this state and that there should be a correspondingly large and favorable entropy of dehydration. The fact that the net entropy change for the U-to-I transition is small implies that this contribution is offset against conformational entropy and, consequently, the structure must have become much more ordered. Inspection of the folded structure of CD2.d1, for instance, shows that there are about 30 buried hydrocarbon residues which would yield around -200 kcal mol^{-1} upon exclusion from the solvent at 25 °C. The data show that half of this entropic contribution is realized in the I-state, a sufficient entropic drive to maintain 75% of the residues in the native backbone conformation.

Transition-State Barrier. The temperature dependencies of the free energy changes associated with the rate-limiting I-to-t ($\Delta G^{\ddagger}_{t-I}$) and t-to-F ($\Delta G^{\ddagger}_{F-t}$) transition barriers for N-PGK and CD2.d1 are shown in panels a and b of Figure 5, respectively. To relate an individual rate constant to the free energy change of activation ($\Delta G^{\ddagger}$) we use the Eyring equation; $k = k'_{(T)} \exp(\Delta G^{\ddagger}_{(T)}/RT)$, where R is the gas constant and $k'_{(T)}$ is the rate of decay of the activated species, i.e., the rate at which this process would occur in the absence of a free energy barrier. While the values for the activation enthalpy ($\Delta H^{\ddagger}$) and the activation heat capacity ($\Delta C_p^{\ddagger}$) do not depend on $k'_{(T)}$, the uncertainty in this value precludes an exact evaluation of the activation entropy ($\Delta S^{\ddagger}$). The values of $\Delta H^{\ddagger}$ and $\Delta C_p^{\ddagger}$ calculated for the I-to-t and t-to-F transitions of N-PGK and CD2.d1 are given in Table 1.

While, for both proteins, formation of the transition state from the intermediate involves further desolvation of the chain (as evidenced by the Δm and ΔC_p values) the lowest energy path for the I-to-F transition at 25 °C has a massive enthalpic barrier (18.2 and 12.6 kcal mol^{-1} for N-PGK and CD2.d1, respectively). Although a large, positive $\Delta H^{\ddagger}$ might be expected for the F-to-t transition, owing to the disruption of van der Waals and hydrogen bonding interactions of the native state, it is, at first sight, more difficult to understand for the I-to-t transition. At least two possibilities exist to explain this phenomenon. First, some have argued that the I-state constitutes a collection of relatively stable, misfolded conformations which must be unfolded before they can achieve the native state (*46*, *47*). Bond breaking in the folding process could, in consequence, provide a positive enthalpy in the transition state. This seems unlikely on structural grounds, however, since the I-states of N-PGK and CD2.d1 have patterns of backbone hydrogen bonding consistent with that in their respective folded states (*25*, *26*). In addition, mutagenesis shows that native contacts are well represented in I-states (*6*−*8*). Nonetheless, it remains possible that there are other, non-native contacts which must be broken to allow progression to the native state.

Second, it may be that the t-state occurs where the protein chain has undergone a critical but necessary stage of dehydration of internal nonpolar groups. In their computer simulations of the formation of small clusters of nonpolar solutes in water, Rank and Baker (*48*) observe a barrier to association occurring at the largest solute separation where the solvent molecules cannot penetrate substantially (*48*). In other words, the barrier to the close-packed association of nonpolar groups (folding from the partially solvated I-state to the fully desolvated F-state) reflects the loss of solute−solvent interactions (dehydration) at a solute separation far from the optimal van der Waals distance.

With regard to the relative entropies of ground and transition states and the evaluation of $k'_{(T)}$, the following arguments are salient. Evidence from measurements of heat capacities and *m*-values shows that the transition state is partially solvated so that its decay to the folded state demands the exclusion of water molecules from the nonpolar core of the protein. If one were to consider a classical reaction in which covalent bonds were being broken and/or made, the rate of decay of the activated complex takes a maximum value of $\sim 10^{13}$ s^{-1} at 25 °C (i.e., k_bT/h), which equates with the frequency of a single bond vibration in the gas phase. Given the requirement for desolvation in the case of a folding reaction (or solvation in the unfolding direction) this value is unrealistically large. To get some idea of $k'_{(T)}$ for a decay which requires movement of water molecules from the activated state we need an estimate of the rate at which this can occur. The Einstein−Smoluchowski equation ($D = d^2/2\tau$) relates the diffusion coefficient (D) of a molecule in a given medium with its diameter (d) and its jump-time (τ), i.e., the time taken for the molecule to diffuse one diameter from its starting position. This produces an average jump-time for a water molecule in *n*-butanol (an approximate mimic of the interior of a partially dehydrated protein) at 25 °C of only $\sim 10^{-10}$ s (value for D taken from ref *49*).

An empirical estimate for $k'_{(T)}$ can be derived from the plot shown in Figure 6. Here the fractional solvent exposure of the transition state ($\alpha^{\ddagger} = m_t/m_U = \Delta C_p^{\ddagger}{}_{F-t}/\Delta C_{pF-U}$) is plotted against the logarithm of the unfolding rate ($\ln k_u$) for a number of different globular proteins (see Analytical Procedures). As can be seen, the correlation between $\alpha^{\ddagger}$ and $\ln k_u$ is remarkably good (correlation coefficient = 0.84), suggesting that the barrier to unfolding depends, at least in part, on the proportion of nonpolar groups which must be solvated before the chain unfolds. From this plot the rate

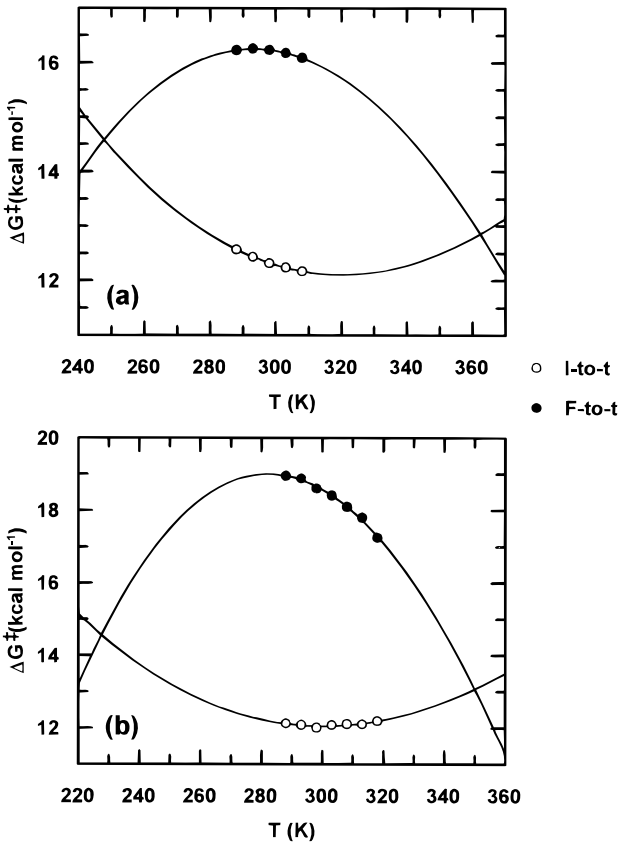

FIGURE 5: Variation of ΔG with temperature: kinetic transitions. The free energy changes ($\Delta G^{\ddagger}$) associated with the I-to-t and F-to-t kinetic barriers as a function of temperature are shown for N− PGK and CD2.d1 in (a) and (b), respectively. For the purposes of these plots, a value for $k'_{(T)}$ at 25 °C of 1×10^{10} s^{-1} was used to calculate $\Delta G^{\ddagger}$ (see Results and Discussion). The data have been fitted to eq 5 to obtain values for the enthalpy, entropy, and heat capacity changes of activation (fits shown as continuous curves). These values are given in Table 1. The activation entropies have also been calculated using a value for $k'_{(T)}$ at 25 °C of 1×10^{7} s^{-1} (see Results and Discussion and Table 1).

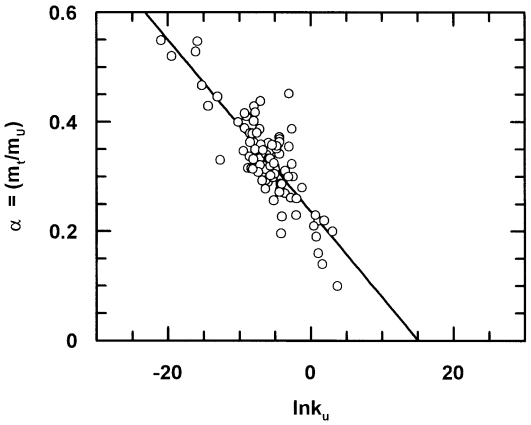

FIGURE 6: Variation of k_u with fractional solvent exposure. In this plot the natural logarithm of the unfolding rate (ln k_u) has been plotted against the fractional solvent exposure of the transition state ($\alpha^{\ddagger} = m_t/m_U$) for a number of different globular proteins at 25 °C (see Analytical Procedures). The data fit to a straight line with a correlation coefficient of 0.84. The intercept, where $\alpha^{\ddagger} = 0$, provides an estimate for $k'_{(T)}$ at 25 °C of $\sim 1 \times 10^{7}$ s^{-1} (see Results and Discussion).

of unfolding extrapolated to a value of m_t/m_U of zero gives a value for $k'_{(T)}$ at 25 °C of $10^{6}-10^{7}$ s^{-1}. This represents a condition in which the transition state and the folded state

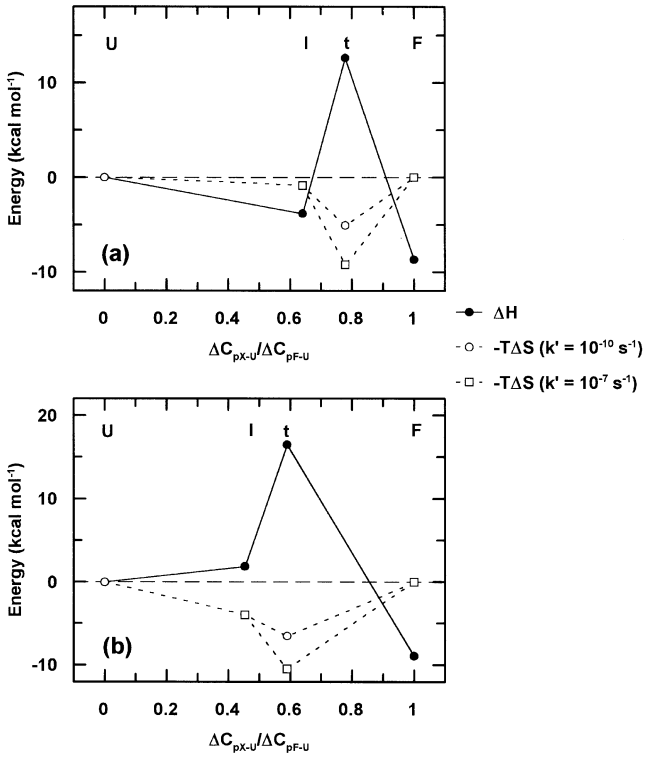

FIGURE 7: Thermodynamic reaction profiles. In (a) and (b), respectively, the enthalpy (ΔH) and entropy ($-T\Delta S$) components of the free energy changes associated with the U-to-I, I-to-t, and t-to-F transitions are plotted against the fractional solvent exposures of the states, relative to F ($\Delta C_{pX-U}/\Delta C_{pF-U}$), for N-PGK and CD2.d1 at their temperatures of maximal stability ($T_{max} = T_o/\exp(\Delta S_{T_o}/\Delta C_p)$; these are 31 and 17 °C for N-PGK and CD2.d1, respectively). The activation entropies calculated using a value of $k'_{(T)}$ of 1×10^{7} and 1×10^{10} s^{-1} are shown, as indicated.

have the same level of solvation, so it can be argued that this marks the upper rate for a structural transition which lacks a barrier. It is interesting to note that this value is in keeping with estimates of $k'_{(T)}$ at 25 °C for protein folding reactions ($\sim 10^{6}$ s^{-1}) made by Hagen et al. (*50*) from a consideration of the upper rate of the rearrangement of a protein chain in aqueous solution.

For N-PGK and CD2.d1 we have calculated $\Delta S^{\ddagger}$ for the I-to-t and t-to-F transitions using values for $k'_{(T)}$ at 25 °C of 10^{7} and 10^{10} s^{-1}. These values are given in Table 1. For both proteins, the values for the I-to-t activation entropies at 25 °C are positive. As the reduction in heat capacity shows the t-state to be less solvated than the intermediate, the origin of this favorable entropy gain must lie in the process of water exclusion from hydrophobic surfaces. Only after the transition state is surmounted is there a net entropic penalty arising from the widespread ordering of side chains in the fully folded state.

The correlation between Δm and solvent exclusion has, in the absence of any other experimental candidate, led to its use as a marker for the reaction coordinate of folding processes (*8*, *23*, *51*). Since ΔC_p and Δm are also strongly correlated it seems reasonable (or at least convenient) to use the ΔC_p value as an indicator of "foldedness" in the representation of thermodynamic data. The data are plotted in this form in Figure 7 at temperatures where the overall entropy change for the folding reaction (ΔS_{F-U}) is zero, i.e., at their respective optimum temperatures. These profiles emphasize the qualitative similarity in the folding behavior

of these topologically dissimilar structures and imply a common set of underlying principles governing the energetics of the folding reaction.

Structural Implications for the Folding Pathway. The experimental results described above are consistent with the following structural model of the principal stages in a folding reaction. As unfolded chains rapidly condense to form the I-state population a large proportion of core residues are desolvated, and the accompanying entropic advantage (with respect to the microscopic free energy of an average chain) is balanced by the entropic cost of narrowing down the conformational freedom of the ensemble. Hence, a high degree of structural order must be attained in this rapid transition; otherwise, if desolvation occurred in a random collapse, the ΔS term would be very large and positive. Side chain orientation is unfixed in this population but the backbone has become topologically defined, largely by crude hydrophobic interactions between core side chains and by intramolecular hydrogen bonding between segments removed from the solvent. To reach the folded state, which is characterized by the orientationally specific docking of side chains, the core must completely dehydrate. The fact that dehydration precedes precise packing is shown by the higher entropy and enthalpy of the transition state. The former is explained by the removal of solvent from the core and the latter by poor van der Waals interactions between side chains and unsatisfied solvent−solvent bonding potentials at this critical stage of solvent exclusion. As the transition state decays to the folded structure, the increasing order of the system ($-T\Delta S$ is highly positive) is driven by bonding contributions from optimal packing and the formation of new solvent−solvent interactions.

REFERENCES

1. Roder, H., and Colón, W. (1997) *Curr. Opin. Struct. Biol. 7*, 15−28.
2. Clarke, A. R., and Waltho, J. P. (1997) *Curr. Opin. Biotech. 8*, 400−410.
3. Kuwajima, K. (1989) *Proteins: Struct., Funct., Genet. 6*, 87−103.
4. Roder, H., ElÖve, G. A., and Englander, S. W. (1988) *Nature 335*, 700−704.
5. Udgaonkar, J. B., and Baldwin, R. L. (1988) *Nature 335*, 694−699.
6. Matouschek, A., Kellis, J. T., Serrano, L., Bycroft, M., and Fersht, A. R. (1990) *Nature 346*, 440−445.
7. Khorasanizadeh, S., Peters, I. D., and Roder, H. (1996) *Nat. Struct. Biol. 3*, 193−205.
8. Parker, M. J., Sessions, R. B., Badcoe, I. G., and Clarke, A. R. (1996) *Folding Des. 1*, 145−156.
9. Fersht, A. R., Serrano, L., and Matouschek, A. (1992) *J. Mol. Biol. 224*, 771−782.
10. Fersht, A. R. (1995) *Curr. Opin. Struct. Biol. 5*, 79−84.
11. Makhatadze, G. I., and Privalov, P. L. (1990) *J. Mol. Biol. 213*, 375−384.
12. Makhatadze, G. I., and Privalov, P. L. (1993) *J. Mol. Biol. 232*, 639−659.
13. Privalov, P. L., and Makhatadze, G. I. (1993) *J. Mol. Biol. 232*, 660−679.
14. Griko, Y. V., Makhatadze, G. I., Privalov, P. L., and Hartley, R. W. (1994) *Protein Sci. 3*, 669−676.
15. Chen, B. L., Baase, W. A., and Schellman, J. A. (1989) *Biochemistry 28*, 691−699.
16. Chen, X., and Matthews, C. R. (1994) *Biochemistry 33*, 6356−6362.
17. Oliveberg, M., and Fersht, A. R. (1996) *Biochemistry 35*, 2738−2749.
18. Schindler, T., and Schmid, F. X. (1996) *Biochemistry 35*, 16833−16842.
19. Segawa, S., and Sugihara, M. (1984) *Biopolymers 23*, 2473−2488.
20. Tan, Y. J., Oliveberg, M., and Fersht, A. R. (1996) *J. Mol. Biol. 264*, 377−389.
21. Hosszu, L. L. P., Craven, C. J., Spencer, J., Parker, M. J., Clarke, A. R., Kelly, M., and Waltho, J. P. (1997) *Biochemistry 36*, 333−340.
22. Driscoll, P. C., Cyster, J. G., Campbell, I. D., and Williams, A. F. (1991) *Nature 353*, 762−765.
23. Parker, M. J., Spencer, J., and Clarke, A. R. (1995) *J. Mol. Biol. 253*, 771−786.
24. Parker, M. J., and Clarke, A. R. (1997) *Biochemistry 36*, 5786−5794.
25. Hosszu, L. L. P., Craven, C. J., Parker, M. J., Lorch, M., Spencer, J., Clarke, A. R., and Waltho, J. P. (1997) *Nature Struct. Biol. 4*, 801−804.
26. Parker, M. J., Dempsey, C. E., Lorch, M., and Clarke, A. R. (1997) *Biochemistry 36*, 13396−13405.
27. Itzhaki, L. S., Otzen, D. E., and Fersht, A. R. (1995) *J. Mol. Biol. 254*, 260−288.
28. Jackson, S. E., el Masry, N., and Fersht, A. R. (1993) *Biochemistry 32*, 11270−11278.
29. Serrano, L., Matouschek, A., and Fersht, A. R. (1992) *J. Mol. Biol. 224*, 805−818.
30. Serrano, L., Kellis, J. T., Cann, P., and Matouschek, A. (1992) *J. Mol. Biol. 224*, 783−804.
31. Hernàndez, E. L., and Serrano, L. (1996) *Folding Des. 1*, 43−55.
32. Yamasaki, K., Ogasahura, K., Yutani, K., Oobatake, M., and Kanaya, S. (1995) *Biochemistry 34*, 16552−16562.
33. Milla, M. E., Brown, B. M., Waldburger, C. D., and Sauer, R. T. (1995) *Biochemistry 34*, 13914−13919.
34. Johnson, C. M., and Fersht, A. R. (1995) *Biochemistry 34*, 6795−6804.
35. Shortle, D., Meeker, A. K., and Freire, E. (1988) *Biochemistry 27*, 4761−4768.
36. Staniforth, R. A., Burston, S. G., Smith, C. J., Jackson, G. S., Badcoe, I. G., Atkinson, T., Holbrook, J. J., and Clarke, A. R. (1993) *Biochemistry 32*, 3842−3851.
37. Myers, J. K., Pace, C. N., and Scholtz, J. M. (1995) *Protein Sci. 4*, 2138−2148.
38. Schellman, J. A. (1955) *Trav. Lab. Carlsberg, Ser. Chim. 29*, 223−229.
39. Frank, H. S., and Evans, M. W. (1945) *J. Chem. Phys. 13*, 507−515.
40. Gill, S. J., Dec, S. F., Olofsson, G., and Wadso, I. (1985) *J. Phys. Chem. 89*, 3758−3765.
41. Muller, N. (1990) *Acc. Chem. Res. 23*, 23−28.
42. Nemethy, G., and Scheraga, H. A. (1962) *J. Chem. Phys. 36*, 3401−3408.
43. Spolar, R. S., Livingstone, J. R., and Record, M. T., Jr. (1992) *Biochemistry 31*, 3947−3955.
44. Tanford, C. (1970) *Adv. Protein Chem. 24*, 1−95.
45. Alonso, D. O. V., and Dill, K. A. (1991) *Biochemistry 30*, 5974−5985.
46. Kiefhaber, T. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 9029−9033.
47. Mirny, L. A., Abkevich, V., and Shakhnovich, E. I. (1996) *Folding Des. 1*, 103−116.
48. Rank, J. A., and Baker, D. (1997) *Protein Sci. 6*, 347−354.
49. Weast, R. C. (1989) *Handbook of Chemistry and Physics*, 70th ed., CRC Press, Boca Raton, FL.
50. Hagen, S. J., Hofrichter, J., Szabo, A., and Eaton, W. A. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 11615−11617.
51. Matouschek, A., and Fersht, A. R. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 7814−7818.

BI972459K