Articles

The Energetics and Cooperativity of Protein Folding: A Simple Experimental Analysis Based upon the Solvation of Internal Residues†

Rosemary A. Staniforth,*,‡ Steven G. Burston,‡ Corinne J. Smith,‡ Graham S. Jackson,‡ Ian G. Badcoe,‡ Tony Atkinson,§ J. John Holbrook,‡ and Anthony R. Clarke‡

Molecular Recognition Centre and Department of Biochemistry, School of Medical Sciences, University of Bristol, University Walk, Bristol BS8 1TD, U.K., and Microbial Technology Laboratory, Public Health Laboratory Service, Centre for Applied Microbiology and Research, Salisbury, Wiltshire SP4 0JG, U.K.

Received August 24, 1992; Revised Manuscript Received January 21, 1993

ABSTRACT: The reversible unfolding of two dissimilar proteins, phosphoglycerate kinase from *Bacillus stearothermophilus* (PGK) and *Staphylococcus aureus* nuclease (SAN), was induced with two denaturants, urea and guanidinium chloride (GuHCl). For each protein, structural transitions were monitored by intrinsic fluorescence intensity changes arising from a unique tryptophan residue. In the case of SAN the single, native tryptophan residue was used, whereas for PGK two versions, one with a tryptophan at position 315 and one at 379, were constructed genetically. The resultant folding curves were analyzed by considering the change in the solvation free energy of internal amino acid residues as the denaturant concentration was varied. We derive the following simple relationship: $-RT \ln K = \Delta G_w + n\Delta G_{s,m}[D]/(K_{den.} + [D])$ where K is the equilibrium constant describing the distribution of folded and unfolded forms at a given denaturant concentration [D], ΔG_w is the free energy change for the transition in the absence of denaturant, and n is the number of internal side chains becoming exposed. $\Delta G_{s,m}$ and $K_{den.}$ are constants derived empirically from the solvation energies of model compounds and represent the behavior of an average internal side chain between 0 and 6 M GuHCl and 0 and 8 M urea. For proteins of known structure these values can easily be derived, and for others, average values in guanidinium chloride ($\Delta G_{s,m}$ = 0.775 kcal/mol and $K_{den.}$ = 5.4 M) or urea ($\Delta G_{s,m}$ = 1.198 kcal/mol and $K_{den.}$ = 25.25 M) can be used in the analysis. Results show that the parameters n and ΔG_w are independent of the denaturant used for all 12 transitions studied. This supports the hypothesis that the unfolding activity of urea and GuHCl can be accounted for by their effect on the solvation energy of amino acid side chains which are buried in the folded but exposed in the unfolded protein. This simple analytical treatment allows the "cooperativity" of protein folding to be interpreted in terms of the number of side chains becoming exposed to the solvent in a given step and allows accurate estimation of the free energy irrespective of the denaturant concentration needed to induce the transition.

An experimentally based description of the process of protein folding demands an understanding of both the structural and energetic properties of intermediates; that is, we need to examine the conformations which exist between the fully unfolded polymer and the condensed native form. Such intermediates arise only transiently when a protein folds in its native solvent and are consequently difficult to study.

One way of stabilizing intermediate structures is to alter the solvent conditions so that they become the predominant species. For these species to be worthy of study, it must then be shown that they not only exist at equilibrium but also arise on the kinetic folding pathway.

The first stage in this process is to identify a solvent which promotes reversible unfolding and determine the relationship between the solvent conditions and the concentration of intermediate conformational states in a structurally defined protein. The next stage is to describe the energetic and structural properties of these states, the latter requiring a knowledge of the physical effect of the chosen solvent.

The most effective and commonly used solvent denaturants for studying the equilibrium folding profiles of proteins are aqueous solutions of guanidinium chloride (GuHCl) and urea. In this paper we attempt to establish, on experimental grounds, the physical relationships between denaturant concentration, the propensity of the medium to solvate the protein interior, and the energetics of a folding transition. Once the principle is demonstrated, we then use it in the general analysis and interpretation of folding profiles in these solvents.

As a starting point for all such analyses, the free energy of folding between two differently folded states of a protein at a given concentration of denaturant, ΔG_d, can be defined:

$$\Delta G_d = -RT \ln (F/U) \quad (1)$$

where U is the concentration of protein in the more denatured state and F the concentration of protein in the more folded state.

The concentrations of F and U can be deduced from the measurement of properties which discriminate between these states, such as intrinsic fluorescence intensity, circular dichroism, or molecular volume.

In order to determine the free energy difference between these species in water, a much more physically relevant property, values for ΔG_d are calculated for the narrow range

† This work was supported by the Science and Engineering Research Council, U.K., and Porton Industries Ltd.

* Author to whom correspondence should be addressed.
‡ University of Bristol.
§ Centre for Applied Microbiology and Research.

0006-2960/93/0432-3842$04.00/0

of denaturant concentrations over which the transition occurs and extrapolated to give a value in the absence of denaturant (ΔG_w). Customarily, this is calculated by linear extrapolation from a plot of ΔG_d versus denaturant concentration, such that

$$\Delta G_d = \Delta G_w + m[\text{denaturant}] \quad (2)$$

There is no physical justification for this method, but it appears that the results obtained from it are reasonably consistent with those obtained using more complex methods (Schellman, 1978). However, such plots become nonlinear at high concentrations of denaturant (Schrier & Schrier, 1977; Aune & Tanford, 1969). The problem therefore arises as to whether this type of linear extrapolation can be made over the full range of denaturant conditions in which folding transitions occur; i.e., will the free energy change of the structural rearrangement in water be correctly estimated in all cases? A second problem lies in interpreting the slope, m, of such a plot. This essentially reflects the "cooperativity" or steepness of the transition and has been interpreted as a measure of the hydrophobicity of a protein (Rose et al., 1985) or the extent of unfolding in the more denatured state (Dill & Shortle, 1991). In the absence of a physical basis for the action of denaturants, a quantitative treatment of the "cooperativity" of a transition is not possible. It should be noted that the term "cooperativity", as applied here to folding profiles, is used to describe the sensitivity of the transition to denaturant concentration. It is not used to mean the degree to which the transition approximates to a two-state model—a definition often applied in protein folding.

The aim of the experiments described is to establish a basis for the ability of two widely studied denaturants, urea and GuHCl, to unfold proteins. From this, we suggest a simple technique for the analysis of equilibrium folding curves which provides a reliable estimation of the free energy of a transition in water and yields a quantitative description of its "cooperativity".

We approach this question by analyzing the unfolding of proteins through a consideration of the known increase in solvation energies of side chains as denaturant concentrations are increased (Nozaki & Tanford, 1963, 1970; Wetlaufer et al., 1964; Pace, 1975). For this study we chose to use intrinsic tryptophan fluorescence as a sensitive and convenient reporter for conformational transitions and to apply the analysis to two structurally diverse model proteins, *Bacillus stearothermophilus* phosphoglycerate kinase (PGK) and *Staphylococcus aureus* nuclease (SAN). The SAN protein has only a single tryptophan residue in its native sequence, and two engineered versions of PGK were used, each of which contained only one tryptophan. As a result, the signal is simplified as it arises from a single reporter group in each case, and the behavior of this signal can be compared when unfolding the proteins in two different denaturants. In addition, both proteins are monomeric, thus avoiding the complexity of quaternary transitions, and their folding transitions are fully reversible.

EXPERIMENTAL PROCEDURES

Production of PGK Molecules Bearing a Single Tryptophan Residue. The *Hin*dIII fragment containing the *B. stearothermophilus* PGK gene (Davies et al., 1991) was subcloned into the phagemid vector Bluescript KS M13(Stratagene, San Diego) in order to facilitate easy production of single-stranded template DNA. This construct was used to transform *Escherichia coli* TG2 cells, and the resultant cell line synthesizes the recombinant PGK as 20–25% of its soluble protein, expression being driven by the endogenous promoter. Single-stranded DNA was prepared as described in the Stratagene documentation.

B. stearothermophilus PGK contains two tryptophan residues at positions 290 and 315 in the protein sequence. To obtain a protein with a single indole fluorophore at position 315, the tryptophan at position 290 was substituted with a tyrosine. A tryptophan-less framework was created by further substitution of the tryptophan at position 315 and then used subsequently for the insertion of a single tryptophan residue at position 379 in place of a phenylalanine. Therefore, two mutants were created, each containing a single indole fluorescent probe.

Site-directed mutagenesis was performed essentially as described by Carter et al. (1985). The host for the site-directed mutagenesis step was the *E. coli* WK6mutS [(lac-proAB), galE, strA/F'lacL^1, ZDM15, proA$^+$B$^+$, mutS::Tn10] cell line. Positive colonies were screened for using radiolabeled oligonucleotide probes as described by Sambrook et al. (1989).

Cell lines were checked for overexpression of PGK using SDS–PAGE, the level of expression being equivalent to that of the wild-type enzyme. Single-stranded DNA from the resultant cell line was generated as described above and the entire gene sequenced to establish that only the directed mutation had arisen.

Oligonucleotides used for mutagenesis and sequencing were synthesized on a Du Pont Coder 300 DNA synthesizer. Sequencing of the mutant genes was performed by the chain-termination procedure of Sanger et al. (1980), using a Du Pont Genesis 2000 automated sequencer. Cells containing the Bluescript plasmid with the PGK insert were selected for using growth media containing ampicillin at 0.5 mg/mL of growth medium.

Purification of Mutant PGK from E. coli Cultures. Cells were grown at 37 °C overnight in NZYCM broth containing ampicillin and harvested by centrifugation, and the cell pellet was resuspended in 50 mM triethanolamine (TEA) (Boehringer Mannheim), pH 7.2, and lysed by sonication. Cell debris was removed by centrifugation, the cell lysate incubated at 55 °C for 15 min in order to select for the thermophilic protein, and the precipitate removed by centrifugation. Ammonium sulfate was then added to 50% saturation to precipitate further contaminants. These were removed by centrifugation, and the remaining mixture was precipitated by increasing ammonium sulfate saturation to 80%. The protein was purified using gel filtration chromatography on a Sepharose CL-6B column and further purified using ion-exchange chromatography on a fast-flow Q-Sepharose column (Pharmacia). PGK eluted at approximately 0.15 M NaCl during a 0–0.3 M NaCl gradient at pH 7.2. The samples gave a single band of apparent M_r 43 000 on an SDS–PAGE, and the method yielded approximately 80 mg of PGK/L of *E. coli* culture. The purified protein was stored as a precipitate in an 80% saturated ammonium sulfate solution at 4 °C. The catalytic rate constants for the mutant enzymes were 227 s^{-1} (W315) and 132 s^{-1} (W379) compared to 272 s^{-1} for the wild-type enzyme. Assays were performed at 25 °C in 25 mM TEA, pH 7.2.

S. aureus Nuclease. This enzyme was obtained from Sigma, U.K. (N-3755) and dialyzed prior to use against 250 volumes of 50 mM TEA, pH 7.4.

Equilibrium Unfolding Titrations. Transitions in protein structure during unfolding were detected by continuously recording the change in the fluorescence intensity of the single tryptophan reporter group as the concentration of denaturant was increased. A 10 μM sample of the native protein in 50

mM triethanolamine hydrochloride buffer at pH 7.5 containing 2 mM dithiothreitol was unfolded by continuously adding a 6 M GuHCl (Bethesda Research Laboratories, enzyme grade) or 8 M urea (BRL, analar) solution in the same buffer. The denaturant solution was added from a continuous titrator at a rate of 1.151 μL/min over a period of 320 min in the case of staphylococcal nuclease with GuHCl and urea and PGK with GuHCl, but over a period of 480 min for the titration of PGK with urea.

Fluorescence intensity measurements were performed on an SLM 8000 single photon counting spectrofluorometer using an excitation wavelength of 295 nm and an emission wavelength of 345 nm. The excitation and emission monochromator slit widths were 4 nm, and the integration time per point in the titration was 10 s. For each profile approximately 2000 data points were collected (or 3000 for the urea titration of PGK) and recorded on computer. The large number of points and long integration time allow a high degree of precision in these measurements.

To ensure that the unfolding steps were at equilibrium during the titrations, a further profile was recorded for each at half the speed of denaturant addition. These curves superimposed on the originals (data not shown). The results were adjusted for the change in volume due to the addition of denaturant.

Use of FACSIMILE Numerical Integration Program. Fluorescence data from unfolding profiles were numerically fitted using a Jacobean matrix algorithm (FACSIMILE; Curtis & Sweetenham, 1985). The total fluorescence intensity (I) was defined (for the most complex unfolding profile, i.e., PGK W379) as

$$I = I_1P_1 + I_2P_2 + I_3P_3 + I_4P_4 \quad (3)$$

where I_1, I_2, I_3, and I_4 are the relative fluorescence intensities of the folded protein, the two intermediates, and the unfolded species, and P_1, P_2, P_3, and P_4 their relative concentrations. The equilibrium system was defined as

$$P_1 \overset{K_1}{\rightleftharpoons} P_2 \overset{K_2}{\rightleftharpoons} P_3 \overset{K_3}{\rightleftharpoons} P_4$$

where K_1, K_2, and K_3 are the equilibrium constants at a given concentration of denaturant. Values for these equilibrium constants were calculated according to the equation:

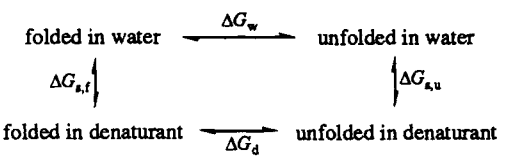
$$K = K_w \exp\{-n\Delta G_{s,m}D/[(K_{den.} + D)RT]\} \quad (4)$$

where K is the equilibrium constant at a given concentration of denaturant D, K_w is the constant in the absence of denaturant, n is the number of side chains solvated in the transition, and $\Delta G_{s,m}$ and $K_{den.}$ are empirically derived solvation constants (see next section). For this system, the program produces a global fit by varying values of K_{w1}, K_{w2}, K_{w3}, n_1, n_2, n_3, I_2, and I_3. The fluorescence of the native protein (I_1) and of the unfolded species (I_4) are directly measured and fixed.

METHODS OF ANALYSIS

Solvation Energies and Protein Stability. The hypothesis that protein unfolding in denaturants can be explained by the increased affinity of internal residues for the solvent was originally formulated by Tanford (1964, 1970; Rowe & Tanford, 1973). In this model it is proposed that unfolding induced by denaturants can be wholly accounted for by an increase in the solvation energy of those parts of the protein which are buried in the native state. Thus, as denaturant is added, the unfolded state becomes more stable because internal residues are no longer driven out of the solvent phase. In energetic terms, this can be represented as shown in Scheme I

Scheme I



I, where $\Delta G_{s,f}$ is the difference in solvation free energy of the folded protein in water and at a defined concentration of denaturant, and $\Delta G_{s,u}$ is the difference in solvation free energy for the unfolded protein. From this:

$$\Delta G_w = \Delta G_d + \Delta G_{s,u} - \Delta G_{s,f} \quad (5)$$

If those residues which become exposed in the unfolded state are identifiable and their increased energy of solvation is known, then

$$\Delta G_{s,u} - \Delta G_{s,f} = \sum N_i\alpha_i\Delta G_{si} \quad (6)$$

where N_i is the number of amino acids of type i in the protein, α_i is the proportion of amino acids of type i which becomes exposed to solvent in a given unfolding step, and ΔG_{si} is the increase in solvation energy of a residue of type i at the given denaturant concentration. From this we have

$$\Delta G_d = \Delta G_w - \sum N_i\alpha_i\Delta G_{si} \quad (7)$$

The values of ΔG_{si} for the different amino acid side-chain types at a given denaturant concentration can be calculated from existing model compound data (Nozaki & Tanford, 1963, 1970; Pace, 1975). Nonetheless, two problems arise. First, the structure of the folded protein must be known, and second, if the protein exhibits more than one transition, assumptions must be made about those parts of the protein which are solvated at each stage of structural change. These are prerequisites for determining the values of N_i and α_i; the first is acceptable in that folding studies are usually performed on proteins with a defined native conformation, and the second defeats inquiry into folding mechanisms since it must presuppose the nature of intermediates.

The Internal Composition of the Model Proteins. To overcome this problem and to produce a simple and more general treatment, the internal amino acid content of both model proteins was analyzed. Table I shows the numbers of each amino acid (N_i) in PGK and SAN and the fraction (α_i) of the surface areas (Connolly, 1983a,b) of all amino acid side chains of each type which becomes exposed to solvent in the unfolding transition from native to a model linear polymer. From these data we can define the composition of the solvent-excluded core of these proteins, i.e., the total number of each type of residue side chain which becomes exposed during unfolding ($n_i = N_i\alpha_i$). As little is known of the structure of the random coil, we have used a fully extended conformation as an approximation. Also the contribution of the backbone atoms to the protein core was not included for reasons developed more fully in the Discussion.

Many previous analyses have taken only aliphatic and aromatic side chains into account despite the fact that more polar residues are often partially or wholly buried (Alonso & Dill, 1991; Shortle et al., 1990). Another approximation has been to consider only amino acids which are more than 95% buried (Rose et al., 1985), although this may also lead to a misrepresentation of what constitutes the protein core.

Relationship between the Solvation Energy of Protein Residues and the Solvent Composition. When proteins are unfolded in increasing concentrations of denaturant, their cores

Table I: An Analysis of the Solvent-Excluded Side Chains[a]

	PGK			SAN			P_{av}		
amino acid	N_i	α_i	n_i	N_i	α_i	n_i	N_i	α_i	n_i
Ala	50	0.630	31.10	13	0.650	8.520	888	0.410	366.7
Cys	1	1.000	1.00	0	0.000	0.000	274	0.650	178.0
Asp	31	0.170	5.25	5	0.440	2.190	625	0.210	128.8
Glu	33	0.255	8.40	11	0.460	5.120	503	0.240	120.7
Phe	15	0.810	12.15	3	0.430	1.310	408	0.680	263.0
Gly	35	0.000	0.00	9	0.000	0.000		0.000	0.0
His	8	0.420	3.40	4	0.210	0.825	317	0.420	133.0
Ile	21	0.875	18.40	5	0.560	2.780	488	0.670	324.5
Lys	32	0.300	10.10	23	0.175	4.050	713	0.175	125.5
Leu	38	0.730	27.70	11	0.750	8.200	795	0.640	508.8
Met	9	0.710	6.40	4	0.560	2.230	142	0.660	93.3
Asn	13	0.330	4.25	6	0.790	4.770	478	0.185	88.4
Pro	15	0.300	4.45	6	0.085	0.510	462	0.230	106.3
Gln	2	0.000	0.00	5	0.330	1.640	375	0.420	157.5
Ser	11	0.720	7.90	3	0.350	1.060	869	0.200	169.4
Thr	13	0.680	8.80	9	0.400	3.590	715	0.290	205.2
Val	38	0.770	29.25	9	0.650	5.850	778	0.630	490.0
Trp	2	0.810	1.62	1	0.680	0.680	161	0.666	107.0
Tyr	8	0.620	4.95	7	0.465	3.250	389	0.550	212.0
n_{total}			191.40			59.000			3875.0

[a] The atomic coordinates of *S. aureus* nuclease (Cotton et al., 1979) and *B. stearothermophilus* PGK were obtained from the Brookhaven Protein Data Bank and from G. J. Davies and H. C. Watson (personal communication), respectively. The fraction of amino acid side chain of type i (α_i) which is buried in the native structure but becomes exposed to solvent when the protein is unfolded was defined as

$$\alpha_i = (A_{ec} - A_f)/A_{ec}$$

where A_f is the summed area of all amino acid side chains of type i exposed to solvent in the folded protein and A_{ec} is the summed area of all amino acid side chains of type i exposed to solvent in an extended chain conformation. A_f and A_{ec} were calculated using the method of Connolly (1983a,b). A_{ec} was obtained by applying the same method to an extended-chain structure of the proteins where $\phi = 180.0°$ and $\psi = 180.0°$ and the side chain is in the gas-phase minimum-energy conformation. This is an approximation that ignores residual structure. This process was repeated for the amino acids of 55 proteins in the Brookhaven Protein Data Bank to determine the internal side-chain composition of an average protein (P_{av}). N_i is the total number of amino acids of type i in the protein(s), and n_i is the number of side chains of type i becoming exposed during complete unfolding.

Table II: Relationship between Change in Solvation Energy and Denaturant Concentration for Protein Side Chains[a]

	ΔG_{si} (cal/mol) at [GuHCl] (M)				ΔG_{si} (cal/mol) at [urea] (M)			
amino acid	1	2	4	6	2	4	6	8
Ala	10	20	30	45	0	−15	−10	−10
Cys[b]	150	245	400	535	115	225	325	415
Asp[c]	200	320	480	645	135	225	330	430
Glu[c]	135	215	315	360	80	130	190	230
Phe	215	355	580	775	180	330	470	600
Gly	0	5	−10	−60	−2	−2	1	0
His	180	285	385	420	100	160	205	255
Ile[b]	135	190	320	430	100	140	205	265
Lys[c]	180	285	385	420	100	160	205	255
Leu	150	210	355	480	110	155	255	295
Met	150	245	400	535	115	225	325	415
Asn	200	320	490	645	135	225	330	430
Pro[b]	100	140	240	320	75	105	155	200
Gln	135	215	315	360	80	130	190	230
Ser[b]	65	90	120	125	40	60	90	115
Thr[c]	65	90	120	125	40	60	90	115
Val[b]	85	115	195	265	60	85	125	160
Trp	400	630	980	1235	270	505	730	920
Tyr	236	386	604	770	225	395	580	735

[a] The table shows the change in solvation energy, ΔG_{si} (cal/mol), of a given side chain when transferred from water to different concentrations of the denaturants urea and GuHCl. The values are derived from solvent partition data and are collected from Nozaki and Tanford [1963 (for urea), 1970 (for GuHCl)]. [b] The values for these amino acid side chains were estimated by Pace (1975) from data collected by Wetlaufer et al. (1964). [c] Those side chains for which no experimental values are available; assumptions about their behavior are described in the text.

become exposed to the solvent. The relationship between the solvation energy of the core residues and the concentration of denaturant must therefore be defined. Values for ΔG_{si}, the change in solvation energy of an amino acid side chain i at a given concentration of denaturant, were taken from Nozaki and Tanford (1963, 1970), Wetlaufer et al. (1964), and Pace (1975). For several amino acids, (glutamate, aspartate, lysine, threonine, and arginine) values of ΔG_{si} are not available. Instead of ignoring them, we assumed that their solvation behavior would closely resemble the pattern of hydration potentials determined by Wolfenden et al. (1981, 1983). Thus aspartate was equated with asparagine, glutamate with glutamine, serine with threonine, and lysine with histidine. Because these side chains contribute little to the composition of the core, errors in these assumptions will have little effect upon the analysis. The contribution of arginine to the core of the average protein is only 2.5% [$\alpha_i = 0.34$, n_i (in P_{av}) = 97 out of a total $n = 3875$]. This, when it is combined with its uniquely high hydration potential (−11.2 kcal/mol) (Von Heijne & Blomberg, 1979) and hydrophilicity (22.3 kcal/

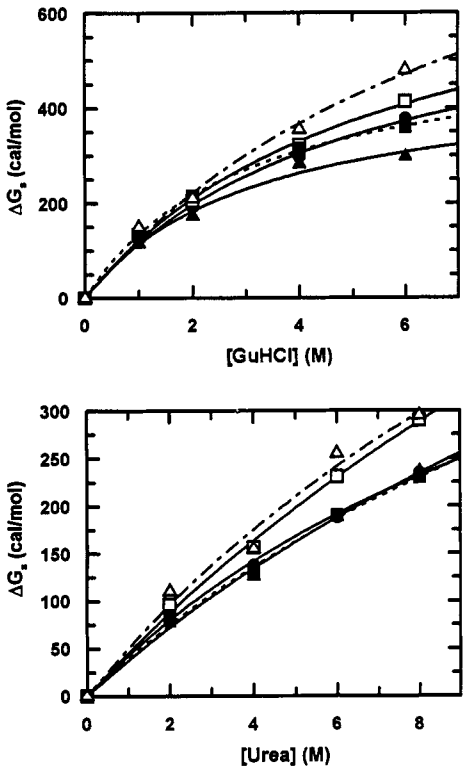


FIGURE 1: Relationship between solvation energy and denaturant concentration for an average internal residue. From the composition of the solvent-excluded core of PGK, SAN, and an "average" protein (P_{av}) (see Table I) the relationship between denaturant concentration and solvation energy (ΔG_s, cal/mol) for an average core side chain was calculated according to eq 9 in Methods of Analysis. Values for both urea and GuHCl are represented by the symbols (●) for SAN, (▲) for PGK, and (□) for P_{av}. As a comparison, the solvation energies of two representative amino acid side chains (dashed lines), glutamate (■) and leucine (△), at different concentrations of the two denaturants are shown. Superimposed are curves which fit the data to the equation:

$$\Delta G_s = \Delta G_{s,m}[D]/(K_{den.} + [D]) \quad (9)$$

to allow determination of the change in solvation energy ΔG_s at any concentration of denaturant [D].

mol) (Wolfenden et al., 1979), implies that ΔG_{si} will be very low for arginine. On these grounds, we excluded arginine from our calculations. Data for the values of ΔG_{si} versus denaturant concentration for both urea and GuHCl are listed in Table II. These values were then used in all subsequent analyses.

Link between the Internal Residue Composition and the Energetics of Protein Folding. Because we can make no assumptions about the structural properties of folding intermediates, we begin by assuming that the internal regions of the protein have a uniform composition of these solvent-excluded residues. The change in solvation energy for an average internal side chain at a given denaturant concentration can then be calculated from a combination of the data from model compounds and the core composition of the protein.

The difference in solvation energy (ΔG_s) between water and a given concentration of denaturant for the average buried side chain is then

$$\Delta G_s = (\sum \Delta G_{si} n_i)/n \quad (8)$$

where ΔG_{si} is the change in solvation energy for a given side chain (i), n_i is the number of side chains of type i in the solvent-excluded core, and n is the total number of side chains in the core. Values for ΔG_s are plotted against denaturant concentration [D] in Figure 1. The data are shown for both PGK and SAN in urea and in GuHCl. It is evident that the relationship between ΔG_s and [D] is not linear. This has been noted previously (Alonso & Dill, 1991; Pace et al., 1990; Shortle & Meeker, 1986). Over the accessible range of denaturant concentration, the data fit well to the empirical, hyperbolic relationship:

$$\Delta G_s(D) = \Delta G_{s,m}[D]/(K_{den.} + [D]) \quad (9)$$

where $\Delta G_s(D)$ is the change in solvation energy for an average core residue between water and the given denaturant concentration and $\Delta G_{s,m}$ is the maximum change in solvation energy at an infinite concentration of denaturant. $K_{den.}$ is a "denaturation constant" measured in molar terms and represents the denaturant concentration required to achieve half $\Delta G_{s,m}$. Values for $\Delta G_{s,m}$ and $K_{den.}$ for both proteins in both denaturants are given in Table IIIA. Santoro and Bolen (1992) have tested the validity of the linear extrapolation method on *E. coli* thioredoxin by deriving values for ΔG_s over a wide concentration range of GuHCl. For this, they took advantage of the fact that, for this protein, thermal denaturation produces similar unfolded states to those occurring at high concentrations of GuHCl or urea. Thermal denaturation of thioredoxin at low concentrations of GuHCl allowed the evaluation of ΔG_s at 25 °C using the Gibbs–Helmholtz equation. They conclude that the relationship between GuHCl concentration and ΔG_s is nonlinear and becomes linear only when a high, but nondenaturing concentration of NaCl equal to that of the denaturant at which the protein unfolds (≥1.5 M) is present in the buffer. Presumably this masks the ionic contribution to the effect of GuHCl; it does not however explain why data for the nonionic denaturant, urea, fit better to a hyperbolic rather than to a linear function (Figure 1).

To obtain constants applicable to the unfolding of an "average" protein in GuHCl or urea, values for $\Delta G_s(D)$ were calculated for an average buried side chain from 55 soluble and globular proteins whose high-resolution coordinates are in the Brookhaven Protein Data Bank and plotted against denaturant concentration. This plot is shown in Figure 1, and the constants are shown in Table IIIA.

If these arguments are correct, then any unfolding transition induced by the denaturants urea or GuHCl can be described by

$$\Delta G_d = \Delta G_w + n\Delta G_s \quad (10)$$

Values for ΔG_s for the given protein and denaturant are calculated from eq 9 using the constants in Table IIIA. ΔG_w is the free energy change for the folding transition in water,

Table III: Denaturation Constants for Internal Residues[a]

	GuHCl		urea	
	$\Delta G_{s,m}$ (kcal/mol)	$K_{den.}$ (M)	$\Delta G_{s,m}$ (kcal/mol)	$K_{den.}$ (M)
(A) Side Chains Only				
phosphoglycerate kinase (PGK)	0.465	3.07	0.922	23.44
staphylococcal nuclease (SAN)	0.677	4.90	0.625	13.60
"average" protein (P_{av})	0.775	5.40	1.198	25.25
(B) Whole Residues				
phosphoglycerate kinase (PGK)	0.859	3.50	0.970	14.16
staphylococcal nuclease (SAN)	1.084	4.43	0.867	11.10
"average" protein (P_{av})	1.099	4.64	1.115	14.90

[a] Figure 1 shows the relationship between solvation energy and denaturant concentration for an average buried side chain in PGK, SAN, and P_{av}. These are fitted empirically to rectangular hyperbolic relationship requiring two constants $\Delta G_{s,m}$ (kcal/mol) and $K_{den.}$ (M) (see eq 13). This table shows the magnitude of these constants calculated (A) for side chains only and (B) for whole residues.

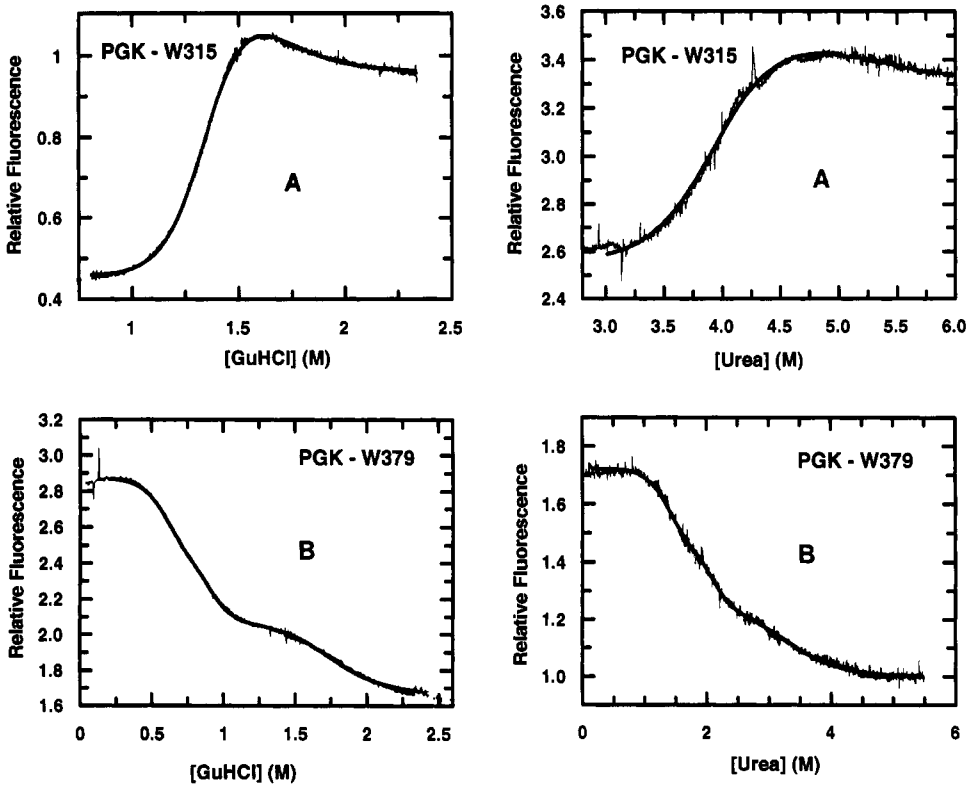


FIGURE 2: Unfolding profiles for PGK. The fluorescence of tryptophan 315 (A) and 379 (B) in PGK was recorded as the protein was unfolded in urea and GuHCl as described in Experimental Procedures. Superimposed on the experimental data are the fitted curves according to eq 13. In the case of W315, two phases are shown, and the curve was fitted to the equilibrium system F ⇌ I ⇌ U (where F is the folded species, I is an intermediate, and U is the unfolded), using the numerical integration program FACSIMILE as described in Experimental Procedures. The optimal fit for W379 was to three transitions (i.e., F ⇌ I_1 ⇌ I_2 ⇌ U) with a residual sum of squares of 27 (the fluorescence range of the data being 120 000). When fitted to two transitions, the residual sum of squares is 58.

and the value n is the number of side chains which become buried in this process. ΔG_d is defined by the equilibrium position of the transition from the less folded (U) to the more folded (F) state, thus:

$$\Delta G_d = -RT \ln([U]/[F]) \quad (11)$$

From the fluorescence intensity data presented in this paper the ratio [U]/[F] is calculated from the relationship:

$$[F]/[U] = (I_0 - I_u)/(I_f - I_u) \quad (12)$$

where I_0 is the recorded fluorescence at a given denaturant concentration and I_u and I_f are the fluorescence intensities of the tryptophan fluorophore in the unfolded and folded states, respectively.

A combination of eqs 9–12 gives the overall equation for the analysis of the fluorescence data:

$$-RT \ln (I_0 - I_u)/(I_f - I_u) = \Delta G_w + n\Delta G_{s,m}[D]/(K_{den.} + [D]) \quad (13)$$

The experimental results presented in the next section are used to evaluate the validity of this analysis and to explore the nature of structural rearrangements in protein folding.

RESULTS

Behavior of Fluorescent Probes in PGK during Folding. Continuous equilibrium unfolding curves for the titration of PGK-315 with both urea and GuHCl are shown in Figure 2A. Two features of these curves are worthy of note. First, as has been observed before (Tanford, 1970; Pace, 1975; Dill & Shortle, 1991), a much higher concentration of urea than GuHCl is required to unfold the protein, and second, the protein does not unfold in a single transition; for both urea and GuHCl an initial large increase in fluorescence is followed by a small reduction. Superimposed on the curves are the fits according to eq 13 using the parameters ($\Delta G_{s,m}$ and $K_{den.}$) determined for PGK shown in Table IIIA. For fitting two overlapping transitions to these experimental curves the numerical integration program FACSIMILE (Curtis & Sweetenham, 1985) was used as described in Experimental Procedures. The result shows that the protein unfolds through an intermediate state (I), thus:

folded (F) ⇌ intermediate (I) ⇌ unfolded (U)

Analysis of the GuHCl result, summarized in Table IVA, shows a free energy change from F to I of −11.4 kcal/mol and an exposure of 77 side chains during the process. The second transition, from I to U, has a ΔG_w of −8.2 kcal/mol and exposes a further 51.5 side chains. When this analysis is compared with that of urea, the result is strikingly similar, the ΔG_w values being −11.25 and −8.15 kcal/mol and the numbers of side chains 83.5 and 52, respectively.

In the case of PGK, we have the advantage of observing the unfolding behavior of the protein from two different probes. One is in the core of the C-terminal domain (residue 315), and the other is also in the C-terminus, but on the interface between the domains (residue 379). The results for the latter probe are shown in Figure 2B; the structure of the protein and the position of these tryptophan residues are represented in Figure 4.

Unfolding as reported from residue 315 shows two distinct transitions, and reference to the atomic structure shows that this probe is likely to be sensitive only to folding events within the C-terminal domain. This is consistent with the C-terminal

Table IV: Summary of Folding Transitions in PGK and SAN[a]

Part A									Part B				
	PGK W315				PGK W379					SAN			
	GuHCl		urea		GuHCl		urea			GuHCl		urea	
parameters	PGK	P_{av}	PGK	P_{av}	PGK	P_{av}	PGK	P_{av}	parameters	SAN	P_{av}	SAN	P_{av}
$\Delta G_{w,1}$ (kcal/mol)					−4.95	−4.10	−3.77	−3.80	$\Delta G_{w,1}$ (kcal/mol)	−4.35	−3.75	−4.21	−4.25
$\Delta G_{w,2}$ (kcal/mol)	−11.40	−9.80	−11.25	−9.84	−10.85	−10.20	−10.78	−10.10	n_1 (side chains)	49.0	41.8	57.3	54.5
$\Delta G_{w,3}$ (kcal/mol)	−8.20	−7.75	−8.15	−7.40	−8.93	−8.20	−8.15	−8.19	$c_{m,1}$ (M)	0.74	0.75	1.80	1.94
n_1 (side chains)					56.7	43.2	55.5	37.6					
n_2 (side chains)	76.8	61.4	83.6	59.6	86.8	65.0	84.8	60.2					
n_3 (side chains)	51.5	42.5	52.2	40.0	51.0	41.1	51.0	42.0					
$c_{m,1}$ (M)					0.71	0.75	1.86	2.30					
$c_{m,2}$ (M)	1.44	1.40	4.00	4.04	1.23	1.37	3.95	4.10					
$c_{m,3}$ (M)	1.60	1.66	4.78	4.60	1.85	1.87	4.90	4.91					

[a] Part A of the table shows the free energy changes (ΔG_w), solvent exposures (n), and transition midpoint values (c_m) for urea- and GuHCl-induced transitions in PGK. This protein shows three transitions $F \rightleftharpoons I_1 \rightleftharpoons I_2 \rightleftharpoons U$. $F \rightleftharpoons I_1$ is the unfolding of the N-terminus, $I_1 \rightleftharpoons I_2$ is the unfolding of the C-terminus to a partially ordered structure, and $I_2 \rightleftharpoons U$ is the complete loss of structure in the N-terminus. W315 reports only events in the C-terminus in the second and third transitions, whereas W379 reports all three (for a description, see the Results Section and Figure 4). Part B of the table shows the results for SAN which shows, essentially, only one transition. As mentioned in the Results, there is evidence of a second transition in GuHCl, the amplitude of which is too small to allow its characterization. By fitting to two transitions, this minor component has been accounted for, to allow the major transition to be analyzed in isolation. In both Parts A and B, the values for ΔG_w, n, and c_m are given for data fitted with two sets of parameters ($K_{den.}$ and $\Delta G_{s,m}$), calculated for the given protein and for an average protein P_{av}. Errors in the n values and for ΔG_w are within ±9% (as calculated for the confidence limits reported by the FACSIMILE algorithm).

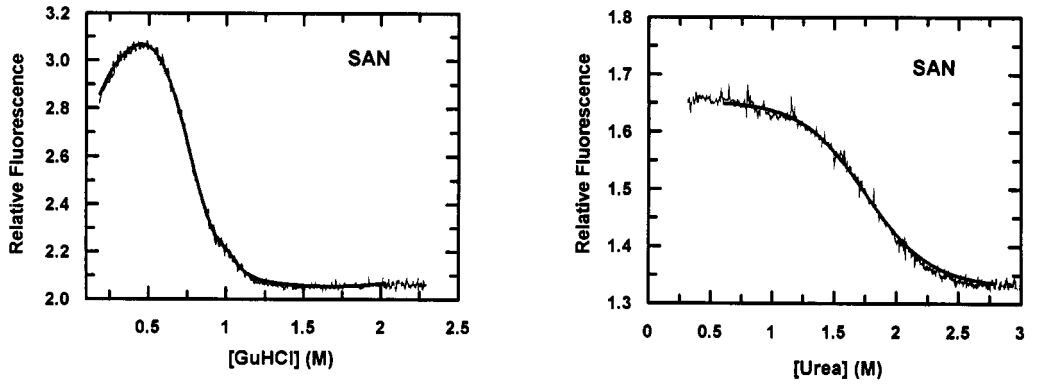


FIGURE 3: Unfolding profiles for SAN. The experiments were performed and the results analyzed as described in the legend to Figure 2. Urea denaturation reveals a single transition, but in GuHCl, there are two minor complications. First, there is a "salt effect" (this is fitted as a simple binding interaction between GuHCl and the protein with a dissociation constant of 51 mM). Such an effect has been seen for other proteins (Pace & Grimsley, 1988). Second, there is a slightly biphasic character to the fluorescence decay in the unfolding step. The magnitude of the second component in this decay is too small to define its properties but has been deconvoluted from the first component in the analysis. The results of the analyses are presented in Table IVB.

domain itself unfolding through an intermediate state. Unfolding reported by residue 379 shows three transitions (see Figure 2B). The second and third of these have properties close to those reported by tryptophan 315; the first represents a new step. The analysis of the free energy changes and side chains exposed in these transitions is again summarized in Table IVA.

Tryptophan 379, being on the interface between the domains, will be sensitive to folding events in the N-terminal domain; the residue is a direct neighbor of the side chains of phenylalanine 146 and leucine 175 and the backbone of glutamate 174. We suggest that the first transition reported by 379 is the unfolding of the N-terminus. Residue 379 appears ideally placed for reporting structural changes across the whole protein molecule, and the sum of all side chains becoming exposed in the three transitions reported by this probe is approximately 193 compared with a theoretical estimate of 191.5 for complete solvation of internal side chains. Other probes, W120, W134, and W169, located in the N-terminus support the assumption that this transition corresponds to a structural change in this domain (S. G. Burston, G. S. Jackson, and A. R. Clarke, unpublished).

Fluorescence Behavior of SAN. The folding curves for SAN are shown in Figure 3, and superimposed are the fits according to eq 13. In urea, the protein is seen to unfold in a single step as expected from many previous investigations (Schecter et al., 1970), and the analysis using parameters calculated for SAN reveals a free energy change of −4.2 kcal/mol and an exposure of 57.3 side chains. This result is very closely reproduced in the case of GuHCl-induced unfolding (see Figure 3 and Table IVB) where the ΔG_w is −4.35 kcal/mol and 49 side chains become exposed. In this case there is evidence of a second transition with a very small amplitude (only 10% of the fluorescence change), but this is so poorly defined that its analysis is statistically unadvisable.

Again, as with PGK, the most striking features are (i) the close correlation between results from each denaturant (the slightly lower values for ΔG_w and n in GuHCl may be explained by the presence of a second, ill-defined transition) and (ii) that the number of side chains becoming exposed during the transitions (49 in GuHCl, 57.3 in urea) is very close to the theoretical number of buried side chains in the folded protein, 59.

DISCUSSION

The analysis of folding curves is an important aspect of protein chemistry for two basic reasons. First, it enables the

free energy of folding of a molecule to be determined, i.e., that energy which, through noncovalent bonding, maintains the native structure over the random coil. This is central to the investigation of protein stability through mutagenic studies, to the design of new proteins, and to the redesign of existing frameworks. Second, it is a central requirement for dissecting folding mechanisms in proteins which fold through intermediate structures [for review see Kim and Baldwin (1990)] because it allows identification and subsequent characterization of these states.

Such an analysis has a long history (Schellman, 1978, 1990; Pace, 1975), and three major methods have been applied. The most commonly used is the linear extrapolation method in which the free energy of folding is plotted against denaturant concentration and the resultant line is extrapolated to give a value in the absence of denaturant (Greene & Pace, 1974). This method assumes no physical mechanism for the action of the denaturant and yields an abstract value for the slope of the line in kcal/mol per molar denaturant (kcal·mol^{-1}·M^{-1}). This value describes the "cooperativity" or steepness of the transition and has been interpreted as a measure of the hydrophobicity of a protein (Pace, 1975); that is, the more hydrophobic a protein, the steeper the transition. It has also been used as a measure of the extent of unfolding in the denatured states (Dill & Shortle, 1991). The experimental analysis presented here effectively combines these two aspects.

A less popular method used to interpret the unfolding curves of proteins is the denaturant binding model which originates from the hypothesis that denaturants bind directly to proteins as ligands during the unfolding process. There is little physical evidence for or against this theory since it is still not demonstrable that attractive interactions between the denaturant and the protein occur more frequently than random collisions (Schellman, 1978). Experimental measurements of denatured binding constants suggest binding is very much weaker than would be expected from model studies (Lee & Timasheff, 1974). Despite these difficulties, unfolding transitions have been fitted to an equation of the form:

$$\Delta G_d = \Delta G_w - \Delta nRT \ln(1 + ka) \qquad (14)$$

where Δn is the difference in the number of denaturant binding sites between folded and unfolded protein, k is an association constant expressing the affinity for these sites, and a is the activity of the denaturant (Tanford, 1970). In these analyses values for k vary from 0.5 to 1 M^{-1} for GuHCl and from 0.08 to 0.3 M^{-1} for urea, depending on the protein (Tanford, 1970; Pace, 1986). Added to the theoretical difficulties, the practical problem with this analysis is that the cooperativity or steepness of a transition is equated, as in the Hill equation, with the number of denaturant binding sites becoming available after the transition. In the original analysis of lysozyme by Tanford (1970) $\Delta n = 7.83$ and $k = 3.00$. Δn appears too small for a protein of 129 residues, and such numbers add little to our interpretation of the underlying structural event. Schellman (1987, 1990) presents a binding model based on selectivity of the protein molecule for a solvent and concludes that the extra energy term involved in interactions with the denaturant (in the potential energy expression of a protein molecule in solution) can be expressed as a power series in the denaturant activity. Empirically, a power series in molarities would approach linearity if there were a Poisson distribution of the denaturant around the protein, or, in the case where the excess free energy is a quadratic function of the denaturant concentration, the distribution would then have a Gaussian character. He suggests that the low values for Δn obtained using the denaturant binding model are due to the denaturant molecules binding as aggregates—this leads to the impossibility of determining the number of binding sites from experimental data.

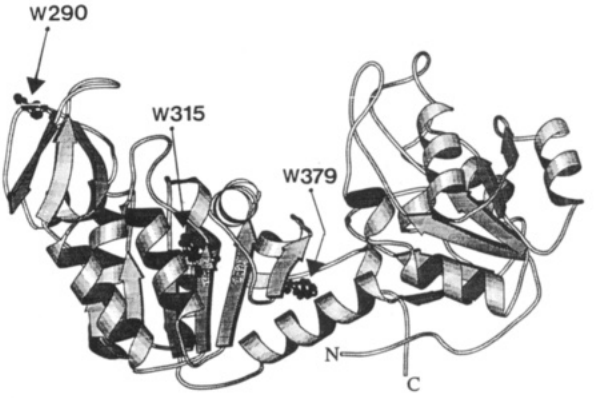


FIGURE 4: Position of tryptophan probes in the PGK molecule. The native structure of *B. stearothermophilus* PGK is represented as a ribbon diagram using the program Molscript (Kraulis, 1991). The wild-type protein contains two tryptophan residues, at positions 290 and 315. For this study the residue at 290 was replaced by tyrosine to create PGK W315; both 290 and 315 were replaced by tyrosine and phenylalanine 379 was replaced by tryptophan to create PGK W379. Positions 290, 315, and 379 are marked on the backbone.

The third approach, which is developed here, is to analyze the folding behavior of proteins in terms of the solvation energy of their component residues. This is an attractive idea on two counts. First, the background information describing the dependence of solvation energy on denaturant concentration for amino acid side chains has been experimentally determined (Nozaki & Tanford, 1963, 1970; Wetlaufer et al., 1964; Pace, 1975). Second, it provides both a testable model for the propensity of proteins to unfold in denaturants and, if the model is valid, secures a means of determining the free energy of folding and of interpreting the shape of folding curves.

For this analysis to be justified, it must be demonstrated experimentally that the unfolding behavior of proteins is primarily dependent upon the change in solvation energy occasioned by denaturant in the medium. Thus if the solvation behavior of urea and GuHCl is correctly represented, the resultant unfolding properties of the protein should be independent of the denaturant used (Tanford, 1970). The results in Table IV show this criterion to be fulfilled; in all cases the free energy of unfolding and the amount of protein core becoming exposed to the solvent for a given transition are independent of the denaturant.

Having justified the treatment, two pieces of information can be derived from its application to folding profiles. First, the free energy change for a given folding step can be estimated. This is also true of both linear extrapolation and denaturant binding models. We find that, for the unfolding of SAN, there is little difference between the value of ΔG_w derived from the linear method (GuHCl, $\Delta G_w = -3.7$ kcal/mol, and urea, $\Delta G_w = -3.6$ kcal/mol) and that derived from the solvation energy method (for GuHCl, $\Delta G_w = -4.35$ kcal/mol, and for urea, $\Delta G_w = -4.2$ kcal/mol). It should be noted that other studies (Shortle & Meeker, 1986), done under different conditions—in 25 mM sodium phosphate and 0.1 M NaCl buffer, pH 7, as opposed to 50 mM triethanolamine hydrochloride, pH 7.5—produced higher values for ΔG_w.

The reason for the correlation of the results between the linear extrapolation and solvation energy treatments is that SAN unfolds at a midpoint denaturant concentration of 1.8 M in urea and 0.7 M in GuHCl. The plots in Figure 1, which

Table V: Results of the Fits to PGK and SAN Folding Transitions Using Denaturation Constants for a "Whole Residue" Treatment[a]

Part A									Part B				
	PGK W315				PGK W379					SAN			
	GuHCl		urea		GuHCl		urea			GuHCl		urea	
parameters	PGK	P_{av}	PGK	P_{av}	PGK	P_{av}	PGK	P_{av}	parameters	SAN	P_{av}	SAN	P_{av}
$\Delta G_{w,1}$ (kcal/mol)					−4.23	−4.30	−3.77	−3.76	$\Delta G_{w,1}$ (kcal/mol)	−3.80	−4.63	−4.24	−5.10
$\Delta G_{w,2}$ (kcal/mol)	−10.57	−9.95	−12.27	−9.9	−10.60	−11.06	−11.08	−10.50	n_1 (side chains)	25.8	30.0	35.6	44.0
$\Delta G_{w,3}$ (kcal/mol)	−8.30	−7.98	−8.20	−7.98	−8.50	−9.10	−8.15	−8.10	$c_{m,1}$ (M)	0.69	0.65	1.77	1.73
n_1 (side chains)					30.2	31.2	33.9	31.6					
n_2 (side chains)	42.4	40.0	57.2	42.2	54.8	56.8	54.9	50.1					
n_3 (side chains)	30.9	29.8	31.5	30.0	28.9	30.0	33.0	30.0					
$c_{m,1}$ (M)					0.68	0.66	1.80	1.78					
$c_{m,2}$ (M)	1.43	1.35	4.02	3.97	1.02	1.00	3.72	3.45					
$c_{m,3}$ (M)	1.59	1.49	5.19	4.66	1.80	1.77	4.84	4.76					

[a] The table shows the free energy changes, ΔG_w (kcal/mol), solvent exposures, n_{wr} (number of whole residues), and transition midpoint values, c_m (M), for urea- and GuHCl-induced transitions in PGK (A) and SAN (B) as calculated from fitting the data to eq 17 using the parameters listed in Table IIIB. Errors in the values for n_{wr} and ΔG_w are within ±9% (as calculated from the confidence limits reported by the FACSIMILE algorithm).

show the relationship between denaturant concentration and average residue solvation energy, demonstrate that over this range the relationship is approximately linear. Thus for proteins undergoing folding transitions in the range 0–1 M GuHCl or 0–2 M urea, the assumption of linearity in the free energy analysis is valid. For larger, stable proteins, such as *B. stearothermophilus* PGK or *B. stearothermophilus* L-lactate dehydrogenase (Smith et al., 1991), where transitions occur in the 1–4 M range of GuHCl this linear analysis would lead to inaccuracies in determining the free energy change during a folding transition.

The second piece of information derived from this analysis is the number of side chains which become exposed in a given unfolding step. This can be taken to represent the cooperativity of the event. For the C-terminus of PGK, probed by tryptophan 315, the sum of both transitions is ~130 side chains (Table IV), close to the theoretical value of ~110. For tryptophan 379, which detects three transitions owing to its interdomain location, the summed value rises to 193. This is close to that estimated for complete solvation of the whole molecule (191.5). Thus the probe at residue 379 which is packed against residues in the N-terminal domain appears to be sensitive to its unfolding and reports this domain to be less stable (see Figure 4). These results are consistent with observations on the folding of PGK drawn from entirely different experimental approaches, which show the N-terminal domain to be less stable than the C-terminal and unfold before it (Freire et al., 1992; S. G. Burston, G. S. Jackson, and A. R. Clarke, unpublished).

A similar close correlation between side chains exposed in unfolding and the number of core side chains is seen with SAN, where the theoretical number is 59 and the experimentally determined number is 49–57.

An important factor to note in this analysis is that we have ignored backbone solvation. This approximation was made due to the lack of consistency in the experimentally derived values for the solvation energy (ΔG_{si}) for peptide groups and their lack of additivity in peptide polymers. Robinson and Jencks (1965) measured the solubilities of polyglycines (di-, tri-, and tetra-) and carbobenzoxy- derivatives of polyglycines at different concentrations of GuHCl and urea. Their conclusions were that, although it seemed that peptide groups contribute to the solvation of proteins by denaturants, the addition of a peptide unit to these model compounds produced changes in solvation energy which depended on the nature of the attached groups. This lack of additivity of ΔG_{si} for a peptide group led us to ignore the contribution of the backbone to ΔG_s in this initial study.

If we were to include the contribution of peptide groups to the solvation of each protein, then:

$$\Delta G_{si(\text{backbone})} = N_i \alpha_i \Delta G_{si(\text{peptide})} \quad (15)$$

where N_i, the number of peptide groups in the protein, and α_i, the fraction of buried peptides, were calculated as described for the side chains and $\Delta G_{si(\text{peptide})}$ was taken from Pace (1975), as calculated from the results of Robinson and Jencks (1965). This contribution can then be added to that of the side chains to give a change in solvation energy for an average whole internal residue:

$$\Delta G_s = \frac{\sum \Delta G_{\text{side chain}} + \Delta G_{\text{backbone}}}{n_{wr}} \quad (16)$$

where n_{wr} is the number of buried whole residues in the folded protein. From this, new values for $\Delta G_{s,m}$ and $K_{den.}$ were derived and the unfolding curves for PGK and SAN were refitted to the new equation:

$$\Delta G_d = \Delta G_w + n_{wr} \Delta G_{s,m}[D]/(K_{den.} + [D]) \quad (17)$$

Values for $\Delta G_{s,m}$ and $K_{den.}$ for PGK, SAN, and P_{av} are listed in Table IIIB. The results of the curve analyses using these new parameters are listed in Table V. There are two important observations to make: (i) Values for ΔG_w have not changed, and (ii) n_{wr} values are very low, suggesting only about 50% of the buried whole residues become exposed on unfolding. Possible reasons for these low numbers are as follows: (1) The contribution made by peptide solvation is overestimated and is unimportant compared to side chains. (2) The "unfolded" protein retains some residual structure (Dill & Shortle, 1991), or random coils are far from fully exposed to solvent (Lee, 1991). (3) The probes used may not report all transitions in the structure of the protein. Therefore, unless reporter groups are inserted at many positions in the protein chain, some steps may remain invisible [in the case of PGK, 14 single tryptophan versions of the protein have been made (S. G. Burston, G. S. Jackson, and A. R. Clarke, unpublished) in order to examine the whole volume of the protein for local structural transitions]. (4) The folded structure is more exposed to solvent than the static Connolly surface analysis predicts. The fact that there are uncertainties in evaluating the contribution of the backbone to the denaturant effects of GuHCl and urea and that consideration of side chains only gives such a good approximation of the internal composition of the protein leads us to favor the latter method. This question, however, requires further consideration.

The analysis used two constants, $K_{den.}$ and $\Delta G_{s,m}$, which are necessary to describe the change in solvation energy for an average internal protein side chain as the denaturant concentration is raised. In Tables IV and V, we show the result of the folding analysis for both SAN and PGK when these constants are derived from the known structure of the individual proteins. Shown for comparison are the results of an analysis where the constants $K_{den.}$ and $\Delta G_{s,m}$ are derived from the solvent-excluded core side chains for 55 proteins in the Brookhaven Protein Data Bank (P_{av}). Thus, global constants are extracted for an average buried side chain in a protein. Whether specifically or globally derived constants are used, there is little difference in the values for the change in free energy and the exposure of residues to solvent in a given transition. This is an important result since it shows that the global constants ($\Delta G_{s,m}$ = 0.775 kcal/mol and $K_{den.}$ = 5.4 M for GuHCl; $\Delta G_{s,m}$ = 1.198 kcal/mol and $K_{den.}$ = 25.25 M) can be conveniently used in the general analysis of equilibrium folding curves without significant loss of accuracy.

ACKNOWLEDGMENT

We thank Drs. H. C. Watson and G. J. Davies for providing us with the cloned *Bacillus stearothermophilus* PGK gene and the high-resolution atomic coordinates, as well as Dr. D. C. Emery for performing the automated DNA sequencing.

REFERENCES

Alonso, D. O. V., & Dill, K. A. (1991) *Biochemistry 30*, 5974–5985.
Aune, K., & Tanford, C. (1969) *Biochemistry 8*, 4586–4590.
Carter, P., Bedouelle, H., & Winter, G. (1985) *Nucleic Acids Res. 13*, 4431–4443.
Connolly, M. L. (1983a) *Science 221*, 709–713.
Connolly, M. L. (1983b) *J. Appl. Crystallogr. 16*, 548–558.
Cotton, F. A., Hazen, E. E., Jr., & Legg, M. J. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 2551–2555.
Curtis, A. R., & Sweetenham, W. P. (1985) United Kingdom Atomic Energy Authority (Unclassified), HMSO Report AERE-R 11771.
Davies, G. J., Littlechild, J. A., Watson, H. C., & Hall, L. (1991) *Gene 109*, 39–45.
Dill, K. A., & Shortle, D. (1991) *Annu. Rev. Biochem. 60*, 795–825.
Freire, E., Murphy, K. P., Sanchez-Ruiz, M. L., & Privalov, P. L. (1992) *Biochemistry 31*, 250–256.
Greene, R. F., & Pace, C. N. (1974) *J. Biol. Chem. 249*, 5388–5393.
Kim, P. S., & Baldwin, R. L. (1990) *Annu. Rev. Biochem. 59*, 631–660.
Kraulis, P. J. (1991) *J. Appl. Crystallogr. 24*, 946–950.
Lee, B. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 5154–5158.
Lee, J. C., & Timasheff, S. N. (1974) *Biochemistry 13*, 257–265.
Nozaki, Y., & Tanford, C. (1963) *J. Biol. Chem. 238*, 4074–4081.
Nozaki, Y., & Tanford, C. (1970) *J. Biol. Chem. 245*, 1648–1652.
Pace, C. N. (1975) *CRC Crit. Rev. Biochem. 3*, 1–43.
Pace, C. N. (1986) *Methods Enzymol. 131*, 266–280.
Pace, C. N., & Grimsley, G. R. (1988) *Biochemistry 27*, 3242–3250.
Pace, C. N., Laurentz, D. V., & Thomson, J. A. (1990) *Biochemistry 29*, 2564–2572.
Robinson, D. R., & Jencks, W. P. (1965) *J. Am. Chem. Soc. 87*, 2462–2470.
Rose, G. D., Geselowitz, A. R., Lesser, G. J., Lee, R. H., & Zehfus, M. (1985) *Science 229*, 834–838.
Rowe, E. S., & Tanford, C. (1973) *Biochemistry 12*, 4822–4827.
Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989) *Molecular Cloning—A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. H., & Roe, B. A. (1980) *J. Mol. Biol. 143*, 161–178.
Santoro, M. M., & Bolen, D. W. (1992) *Biochemistry 31*, 4901–4907.
Schechter, A. N., Chen, R. F., & Anfinsen, C. B. (1970) *Science 167*, 886–888.
Schellman, J. A. (1978) *Biopolymers 17*, 1305–1322.
Schellman, J. A. (1987) *Biopolymers 26*, 549–559.
Schellman, J. A. (1990) *Biopolymers 29*, 215–224.
Schrier, M. Y., & Schrier, E. E. (1977) *J. Chem. Eng. Data 22*, 73–74.
Shortle, D., & Meeker, A. K. (1986) *Proteins: Struct., Funct., Genet. 1*, 81–89.
Shortle, D., Stites, W. E., & Meeker, A. K. (1990) *Biochemistry 29*, 8033–8041.
Smith, C. J., Clarke, A. R., Chia, W. N., Irons, L. I., Atkinson, T., & Holbrook, J. J. (1991) *Biochemistry 30*, 1028–1036.
Tanford, C. (1964) *J. Am. Chem. Soc. 86*, 2050–2059.
Tanford, C. (1970) *Adv. Protein Chem. 24*, 1–95.
Von Heijne, G., & Blomberg, C. (1979) *Eur. J. Biochem. 97*, 175–181.
Wetlaufer, D. B., Malik, S. K., Stoller, L., & Coffin, R. L. (1964) *J. Am. Chem. Soc. 86*, 508–514.
Wolfenden, R. (1983) *Science 222*, 1087–1093.
Wolfenden, R. V., Cullis, P. M., & Southgate, C. C. F. (1979) *Science 206*, 575–577.
Wolfenden, R., Anderson, L., Cullis, P. M., & Southgate, C. B. (1981) *Biochemistry 20*, 849–855.