Folding Kinetics of T4 Lysozyme and Nine Mutants at 12 °C†

Bao-Lu Chen,‡ Walter A. Baase, Hale Nicholson,§ and John A. Schellman*
Institute of Molecular Biology, University of Oregon, Eugene, Oregon 97403-1229
Received August 1, 1991; Revised Manuscript Received November 1, 1991

ABSTRACT: The kinetics of unfolding and refolding of T4 lysozyme and nine of its mutants have been investigated as a function of guanidinium chloride concentration at 12 °C. All show simple two-state, first-order kinetics. Two types of mutants were studied: proline–alanine interchanges and substitutions at position 3 with side chains of varying hydrophobicity. Crystal structures are available for seven of the ten proteins. The effect of mutations on the folding kinetics is more pronounced and complex than on equilibrium thermodynamics. The proteins fall into two broad kinetic classes with one class rather close to the wild type. P86A is a mutant with marked changes in kinetics but only a very small change in stability. Since the 86 position is in the middle of an α-helix, the indications are that the helix containing an A residue is more stable in the transition state than one containing a P residue. The other mutants are more complicated, with the refolding and unfolding rates unequally affected by the mutations. On the basis of comparisons with other investigations, we conclude that the rate-determining step in the presence of guanidinium chloride is not the same as in aqueous solution and that it most likely precedes it. The indications are that we are studying the formation of a transition intermediate which is destabilized by the denaturant and which resembles the A intermediate of the framework or molten globule models for protein folding.

The unfolding of proteins by lowering the temperature is a process which has now been observed many times [see Privalov (1990) for a review]. Proteins do not normally unfold spontaneously at low temperatures, and the reaction is usually observed by imposing a partially destabilizing condition such as extremes of pH, high pressure, or the presence of denaturing agents. Two earlier studies from our laboratory demonstrated that a mutant of T4 lysozyme possessing a 3–97 disulfide bridge undergoes a cold unfolding reaction in the presence of guanidinium chloride (Chen & Schellman, 1989; Chen et al., 1989).

One of the interesting features of the unfolding–refolding reaction at low temperatures in guanidinium chloride is the extreme slowness of the reaction. Half-times can be many minutes or even hours, and it is possible to follow the reaction kinetics using a slow measurement technique like circular dichroism with conventional instrumentation. This phenomenon turns out to be general for all of the T4 lysozymes we have tried, and it was in fact observed for β-lactoglobulin by Pace and Tanford many years ago (Pace & Tanford, 1968).

Careful studies of the forward and backward rates as a function of temperature and guanidinium chloride concentration, as were reported in our earlier studies (Chen et al., 1989), can be very time consuming. For this reason, we decided to do an exploratory survey of a number of proteins using a constant-temperature procedure devised by Matthews and his co-workers (Matthews & Hurle, 1987). This is a method which is applicable to a two-state reversible reaction of the form given in eq 1 below, for which the apparent, first-order rate constant is $k_r = k_1 + k_2$.[1] The procedure is to plot ln k_r vs C, the concentration of denaturant (or equivalently the log of the relaxation time vs C). The plot has the form of a chevron, and from it one can estimate the forward and backward rate constants in the absence of added denaturant and the change in interaction of the protein with the denaturant during the activation process. It is especially well suited for the comparison of the kinetic properties of mutant proteins at a given temperature. The analysis of these plots has been described in several papers (Beasty et al., 1986; Hurle et al., 1987; Matthews & Hurle, 1987; Perry et al., 1989; Chen et al., 1989), and will be discussed in the section on the interpretation of chevron plots. All the measurements are at 12 °C, which is close to the temperature of maximum stability of the proteins in guanidinium chloride solutions.

The methods used permit a quite accurate determination of rates for the unfolding–refolding reaction, which has been demonstrated to be reversible and first order. These rates, however, are very strongly perturbed by the presence of guanidinium chloride, and major questions arise about the relationship between the folding reaction under these experimental conditions and the folding reaction in aqueous solution and about the relation between the transition state observed in the presence of denaturants and the various intermediates which have been observed or proposed for folding under normal

† This work was supported by NIH Grants GM20195 (J.A.S.) and GM21967 (B.W. Matthews) and T32 GM07759 Molecular Biology and Biophysics Training Grant (H.N.) and by NSF Grant PCM8609113 (J.A.S.).

‡ Present address: Department of Pharmaceuticals, Amgen, Inc., Thousand Oaks, CA, 91320.

§ Present address: Department of Chemistry and Biochemistry, Massey University, Palmerston North, New Zealand.

[1] Abbreviations: C, molar concentration of guanidinium chloride; C_m, concentration of guanidinium chloride at the midpoint of the transition; K, equilibrium constant for unfolding (K = [U]/[N]); k_r, apparent first-order rate constant ($k_r = k_1 + k_2$); k_1, rate constant of unfolding (s^{-1}); k_2, rate constant of folding (s^{-1}); k_1°, rate constant of unfolding (s^{-1}) extrapolated to $C = 0$; k_2°, rate constant of folding (s^{-1}) extrapolated to $C = 0$; ΔG°, extrapolated molar Gibbs free energy of unfolding; $\Delta G_1^{\circ\ddagger}$, extrapolated apparent molar free energy of activation for unfolding; $\Delta G_2^{\circ\ddagger}$, extrapolated apparent molar free energy of activation for folding; $\alpha^\ddagger = RT\,\Delta\beta_1^{\circ\ddagger}/RT\,\Delta\beta^\circ$, solvent exposure factor of the transition state relative to the unfolded state; $RT\,\Delta\beta^\circ$, change in molar cosolvation free energy (slope of ΔG vs C); $RT\,\Delta\beta_1^{\circ\ddagger}$, change in molar cosolvation free energy of activation for the unfolding reaction ($-\Delta\beta_1^{\circ\ddagger}$ is the slope of the plot of ln k_1 vs C on the far right of the chevron plot); $RT\,\Delta\beta_2^{\circ\ddagger}$, change in molar cosolvation free energy of activation for the folding reaction ($-\Delta\beta_2^{\circ\ddagger}$ is the slope of the plot of ln k_2 vs C on the far left of the chevron plot); WT, wild type; A, C, F, G, I, L, P, V, and Y, one-letter codes for alanine, cysteine, phenylalanine, glycine, isoleucine, leucine, proline, valine, and tyrosine, respectively. Symbols like A82P indicate the replacement of the wild-type alanine at position 82 by proline.

0006-2960/92/0431-1464$03.00/0

aqueous conditions. These questions are addressed in the final section, and tentative answers are proposed on the basis of extrapolation to zero guanidine concentration and a comparison with the results of other investigations.

Materials and Methods

Mutant T4 lysozymes were constructed by oligonucleotide-directed mutagenesis (Zoller & Smith, 1984). Procedures for mutagenesis, cloning, DNA sequencing, and protein purification are described elsewhere (Muchmore et al., 1989; Alber et al., 1988; Matsumura et al., 1988; Matthews et al., 1987; Alber & Matthews, 1987). The purity of proteins was examined by SDS–polyacrylamide gel electrophoresis and reverse-phase HPLC.

Ultrapure guanidinium chloride was purchased from AMRESCO Co. and was used without further purification. All other reagents were Baker analyzed reagents.

All solutions were made using doubly distilled water filtered through a 0.22-μm Millipore filter. Protein concentration (0.015–0.03 mg/mL) was estimated from optical density at 280 nm using a molar extinction coefficient of 24 170 M^{-1} cm^{-1} (Elwell & Schellman, 1975). Guanidinium chloride concentrations of the samples were calculated from their refractive index (Nozaki, 1972) measured after each experiment. The pH of the samples was measured at room temperature with a Radiometer PHM84 meter with a GK2421C electrode and the Radiometer S series of pH standards.

Circular dichroism (CD) measurements were done in a JASCO 500-C spectropolarimeter. HELLMA suprasil cuvettes (1-cm optical path length) were used. Instrumentation for the controlling of sample temperature and data collection were the same as reported previously (Chen & Schellman, 1989).

The unfolding–refolding experiments made use of concentration-jump methods. For unfolding experiments, the protein stock was mixed with the desired guanidinium chloride solution at 12 °C. For refolding experiments, the protein was first denatured at 4–7 M denaturant and then the solution was diluted to the desired denaturant concentration at the experiment temperature to renature the protein. The solutions were prepared so that the final solution contained 50 mM phosphate buffer and 1.0×10^{-4} M dithiothreitol. The change of CD at 223 nm upon denaturation and renaturation was recorded starting approximately 20 s after mixing. The treatment of these kinetic data was the same as reported elsewhere (Chen et al., 1989). The final plateau values from kinetic runs were used for constructing the equilibrium unfolding curves.

Kinetic data were recorded for at least 10 half-times of the reaction and were fitted to a single exponential using points from 0 to 90% completion with a weighted least-squares subroutine from Bevington (1969).

Chevron Plots and Their Interpretation. Except where it is explicitly stated otherwise, the folding and unfolding reactions of the proteins, followed by measuring CD at 223 nm, are two-state and are expressible in terms of a single relaxation time. Our criteria for making this judgment have been described in a previous paper (Chen et al., 1989). Consequently, we are dealing with the kinetic scheme

$$N \underset{k_2}{\overset{k_1}{\rightleftharpoons}} D \qquad k_c = 1/\tau = k_1 + k_2 \tag{1}$$

where τ and k_r are the first-order relaxation time and rate constant, respectively. The measurements reported in this paper are of k_r as a function of the concentration of guanidinium chloride, C.

In dealing with data of this kind we follow the chevron procedure of C. R. Matthews and his associates (Matthews & Hurle, 1987) discussed above and in the section on the treatment of the data. The main observation of this paper are the changes in these plots produced by substitutional changes in the primary sequence of the protein.

One of the characteristics of these plots is that they become linear at high C (unfolding reaction) and at low C (folding reaction) (see Figure 1). By assuming the validity of transition-state analysis, the logarithm of a rate constant is given by

$$\ln k = \ln \frac{\kappa k_b T}{h} - \frac{\Delta G^{\ddagger}}{RT} \tag{2}$$

where κ is a transmission factor and $\Delta G^{\ddagger}$ is the free energy of activation and the other symbols are standard. It is known that the Eyring formula with $\kappa = 1$ is an upper bound for the rate (Berne et al., 1988). Since many paragraphs in theoretical papers in recent years have been devoted to the deficiencies of transition-state theory, it may as well be admitted that the preexponential factor may not be k_bT/h in the rate equation so that subtracting $\ln (k_bT/h)$ from $\ln k$ may not yield the free energy of the transition state. If it is granted that a transition-state theory with a correct preexponential factor is appropriate, the free energies calculated from the Eyring formula will include a correction term. Because of this caveat, $\Delta G_1^{\circ\ddagger}$ and $\Delta G_2^{\circ\ddagger}$, the free energies of activation extrapolated to $C = 0$ will be called apparent free energies of activation for unfolding and refolding, respectively. They will provide adequate comparisons of the various mutants if the true preexponential factor does not vary from protein to protein. The extrapolation of $\ln k_1$ and $\ln k_2$ to zero denaturant ($C \rightarrow 0$) does not depend on transition-state theory but does assume that linearity persists outside the range of experimental data.

The linear dependence of $\ln k$ on C signifies a linear dependence of $\Delta G^{\ddagger}$ on C and is reminiscent of the linear dependence of $\ln K$ on C, which has been found to be a very good approximation for many proteins (Pace, 1986; Santoro & Bolen, 1988; Bolen & Santoro, 1988). K is the equilibrium constant, given by

$$K = \frac{[U]}{[N]}$$

Solution chemists normally denote excess or transfer free energies by the symbol $RT\beta$, where β is the free energy of transfer in units of RT. In protein denaturation studies, β is very often found to be linear in the concentration of denaturant and then can be conveniently written as $\beta^{\circ}C$, where β° is the free energy per unit molarity and C is the molarity of the denaturant. Note that β is a function of concentration; β° is not. The change in solvation free energy on unfolding is given by $RT\Delta\beta^{\circ}C$, where $\Delta\beta^{\circ} = \beta_U^{\circ} - \beta_N^{\circ}$; β_U° is the free energy of solvation by the denaturant of the unfolded form and β_N° is the equivalent quantity for the folded form. $-RT\Delta\beta^{\circ}$ is often referred to as the "m value" in the denaturation literature following Pace (1975). We prefer the more elaborate symbol with its direct association with solution theory (Schellman, 1978) but will refer to this quantity as the molar "cosolvation free energy" since it is the solvation free energy per mole associated with the addition of a second solvent component, the denaturant. The linear equilibrium relation may then be conveniently expressed in the form $\ln K = \ln K^{\circ} - \Delta\beta^{\circ}C = \ln K^{\circ} + mC/RT$. The correctness of the linear relation is easily checked by plotting $\ln K$ vs C, in which case a straight line is obtained with intercept $\ln K^{\circ}$ and slope $\Delta\beta^{\circ}$ (see Figure 5).

The linearity at the high and low concentration wings of the curves in Figure 1 and subsequent figures indicates that the effect of guanidinium chloride on the kinetics of the unfolding reaction may be interpreted in a similar fashion (Hurle et al., 1987; Matthews & Hurle, 1987; Kuwajima et al., 1989; Chen et al., 1989). The analogous linear expressions for the rate constants are

$$\ln k_1 = \ln k_1^\circ - \Delta\beta_1^{\circ\ddagger} C \quad (3a)$$

$$\ln k_2 = \ln k_2^\circ - \Delta\beta_2^{\circ\ddagger} C \quad (3b)$$

where k_1 and k_2 are the rate constants of the unfolding and folding reactions in the presence of guanidine, k_1° and k_2° are the constants in the absence of denaturant, $\Delta\beta_1^{\circ\ddagger} = \beta^{\circ\ddagger} - \beta_N^\circ$, and $\Delta\beta_2^{\circ\ddagger} = \beta^{\circ\ddagger} - \beta_U^\circ$. $\beta^{\circ\ddagger}$ is the free energy of transfer of the transition state from aqueous solution to denaturant solution per unit of molarity. Thus, $RT\Delta\beta_1^{\circ\ddagger}C$ is the change in activation free energy of the unfolding induced by the presence of the denaturant. It is negative. Similarly, $RT\Delta\beta_2^{\circ\ddagger}C$, which is positive, is the change in activation free energy of the folding reaction induced by the denaturant.

Treatment of Data. Relaxation rate constants, k_r, are obtained from kinetic data as the slopes of first-order logarithmic plots (Chen et al., 1989). $\ln k_r$ is then plotted as a function of C to obtain a V-shaped curve as shown in Figure 1. Assuming that the activation free energy is a linear function of C for single-step, reversible, first-order kinetics, then $k_r = k_1 + k_2$ and

$$\ln k_r = \ln(k_1^\circ e^{-\Delta\beta_1^{\circ\ddagger}C} + k_2^\circ e^{-\Delta\beta_2^{\circ\ddagger}C}) \quad (4)$$

At high C, the unfolding reaction dominates almost completely and eq 4 reduces to eq 3a, showing that the intercept and slope of the linear branch on the right are $\ln k_1^\circ$ and $-\Delta\beta_1^{\circ\ddagger}$, respectively. Similarly, the intercept and slope of the left branch are $\ln k_2^\circ$ and $-\Delta\beta_2^{\circ\ddagger}$. The dashed lines in Figure 1 are plots of eqs 3a and 3b using the values of the parameters determined from the complete chevron plot. C_m, the concentration where $K = 1$ for the reaction, can be obtained as the point of intersection of the two linear regions, where $k_1 = k_2$.

Equation 4 was subjected to nonlinear least-squares analysis with $\Delta G_1^{\circ\ddagger}$, $\Delta G_2^{\circ\ddagger}$, $\Delta\beta_1^{\circ\ddagger}$, and $\Delta\beta_2^{\circ\ddagger}$ as parameters, using the MLAB program (Knott, 1979). $\Delta G_1^{\circ\ddagger}$ and $\Delta G_2^{\circ\ddagger}$ are thereby *defined* by eq 1 with $\kappa = 1$. The values reported for these quantities in Table II are averages of two to four determinations. The accuracy of the parameters for this linear interpretation is enhanced by two factors: (1) The curve fitting must not only match the linear parts of the curve but also represent the vertex in detail; (2) the difference in the intercepts, $\ln k_1 - \ln k_2$, must equal $\ln K$, which is verified independently.

Error Analysis. The nonlinear least-squares program provides estimates of error for all parameters. For the best determined systems (WT and A82P), the random errors in the slopes ($\Delta\beta_1^{\circ\ddagger}$ and $\Delta\beta_2^{\circ\ddagger}$) are about 2%, while the errors in the apparent free energies are less than 1 kcal. For the poorest determined systems (I3F, I3A, and I3G), the errors in the slopes are about 10% while the errors in the apparent free energies of activation are of the order of 2.5 kcal for the folding reaction and about 1.0 kcal for the unfolding reaction. The latter do not include the effect of undetected curvature on the extrapolation or systematic errors. Translating the error in $\Delta G_2^{\circ\ddagger}$ into the error in the folding rate, k_2°, indicates that our estimate may be off by a factor 1.5 for WT and A82P and by factors of 9, 13, and 16, respectively, for I3G, I3A, and I3F. In other words, the estimates of limiting rates at $C = 0$ may err by about 50% for the best experiments and by about an order of magnitude for the least accurate. These estimates will be of importance in the conclusion section where the extrapolations are compared with other results.

Results

The general course of the experiments was to determine the first-order rate constant for the approach to equilibrium as a function of guanidinium chloride concentration and then to analyze the data via chevron plots to obtain the dependence of the logarithms of the forward and backward reactions on the concentration of denaturant (see the preceding section). This simple scheme suffices because the unfolding and refolding of the lysozymes under the conditions of this study are two-state within experimental error as can be seen, for example, from the character of the chevron plots. Other studies on T4 lysozyme under different conditions have yielded more complex kinetics. Desmadril and Yon concluded that kinetic intermediates exist by the comparison of kinetic and equilibrium results obtained by fluorescence and circular dichroism (Desmadril & Yon, 1984). Fink and his collaborators have performed experiments at low temperatures, varying guanidinium concentration in methanol–water mixtures, and have found evidence for intermediates (Anderson et al., 1990). The former have interpreted their results in terms of a sequential folding mechanism, the latter in terms of domain folding.

Wild-Type T4 Lysozyme: Equilibrium and Kinetics. The equilibrium free energy of unfolding of wild-type T4 lysozyme was determined in the usual way by evaluating $\ln K$ from plots of CD_{223} vs guanidinium concentration (Santoro & Bolen, 1988). Because of the slowness of the reactions at low temperature, long times were frequently required for equilibrium to be attained. Our procedures for the establishment of equilibrium conditions have been described in Chen and Schellman (1989). Plots of $-RT \ln K$ vs denaturant concentration were linear within experimental error. The intercept of the least-squares line is taken as an estimate of the free energy of unfolding in the absence of added denaturant (Pace, 1986; Santoro & Bolen, 1988). The slope of the line, $RT\Delta\beta^\circ$, is the change in excess free energy of the unfolding reaction per mole of added denaturant or, equivalently, the change in solvation free energy of unfolding per mole of added denaturant. It provides a measure of the solvation of the groups which are exposed when the protein is unfolded (Schellman, 1978). The results for wild type at pH 5.7 and $T = 12$ °C are $\Delta G^\circ = 17.6$ kcal/mol and $RT\Delta\beta^\circ = -6.12$ kcal/mol M^{-1}. The equilibrium results for the wild type and all the mutants are presented in Table I.

The value of −6.12 kcal/mol found for the cosolvation free energy is considerably greater in magnitude than the value of −4.65 found in a previous publication for a mutant (I3C-C97/C54T) with an engineered S–S bridge at pH 5.0 (Chen & Schellman, 1989). Since these two proteins differ only by two residues out of 164, it is difficult to understand this result if the final states of the unfolding for the two proteins are the same, since the cosolvation free energy is the sum of the contributions of all groups exposed by the unfolding and one residue should not make a disproportionate difference unless structural differences occur in either the native or the unfolded state. We interpret the smaller value of $-RT\Delta\beta^\circ C$ as an indication of the lower exposure to solvent of the unfolded form of the protein with the disulfide mutant. Evidently, the long covalent loop in this molecule holds the side chains in closer proximity and facilitates the retention of interchain contacts in the unfolded state. These contacts may take the form of small hydrophobic clusters.

The chevron plot for wild-type T4 lysozyme is shown in Figure 1. The solid line is the best fit of the data to the model

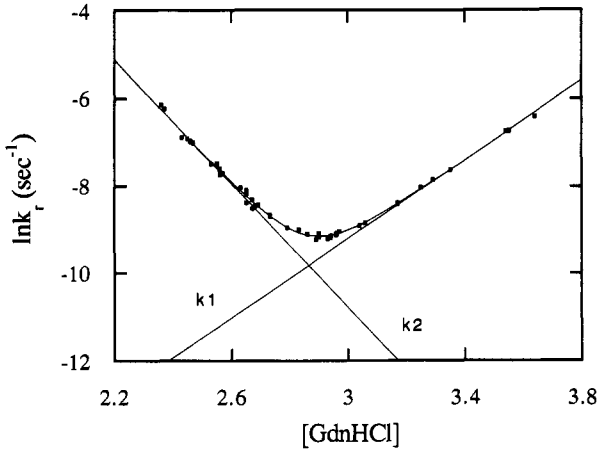

FIGURE 1: ln k_r vs concentration of guanidinium chloride. The points are experimental data for wild-type lysozyme (T = 285 K, pH = 5.7). The curve is the least-squares plot of eq 4 obtained from these data. The parameters of the curve are in Table II. The straight lines are the theoretical plots of k_1 and k_2 based on eqs 3a and 3b using the same parameters. Two points at high concentration are missing from this plot (see Figure 3 and its discussion in the text).

in which ln k_1 and ln k_2 are linear in guanidinium chloride concentration. This is the curve with which all the mutants will be compared.

Table I: Equilibrium Measurements (T = 12 °C, pH = 5.7)

protein	$\Delta G°^a$	$RT\ \Delta\beta°$	C_m	T_m (pH = 2)	$\Delta G°$-(kin)^b	$\Delta G°$-(eq)^b
WT	18.2	−6.34	2.88	41.9	18.8	17.6
A82P	18.6	−6.34	2.93	42.7	18.8	18.0
P37A	17.9	−6.23	2.87	41.1	18.7	17.1
P86A	17.8	−5.93	3.01	40.6	18.4	17.8
L39P	15.4	−6.19	2.49	36.4	15.3	15.5
P143A	17.2	−6.66	2.58	36.9	16.3	17.4
I3V	18.1	−6.50	2.78	39.8		
I3F	16.8	−6.45	2.61	37.9		
I3A	15.7	−6.23	2.52	38.1		
I3G	15.9	−6.82	2.34	34.7		
I3C-C97/C54T^c	14.7	−4.65	3.16	69^d		

^{a} Units: $\Delta G°$, kilocalories per mole; $RT\ \Delta\beta$, kilocalories per mole per molarity unit; C, molarity unit; T, degrees Celsius. ^{b} For a discussion of the equilibrium and kinetic methods of calculating $\Delta G°$, see the section on wild-type T4 lysozyme. ^{c} Experiments done at pH 5.0. ^{d} pH = 6.5.

There are two high-concentration, short-time points for the wild-type protein which are indicated in Figure 3 but were not present in the demonstration Figure 1. We do not know whether these points represent a slower process, which becomes rate-limiting at high concentrations of guanidine, or whether the deviations result from a systematic error. These two points

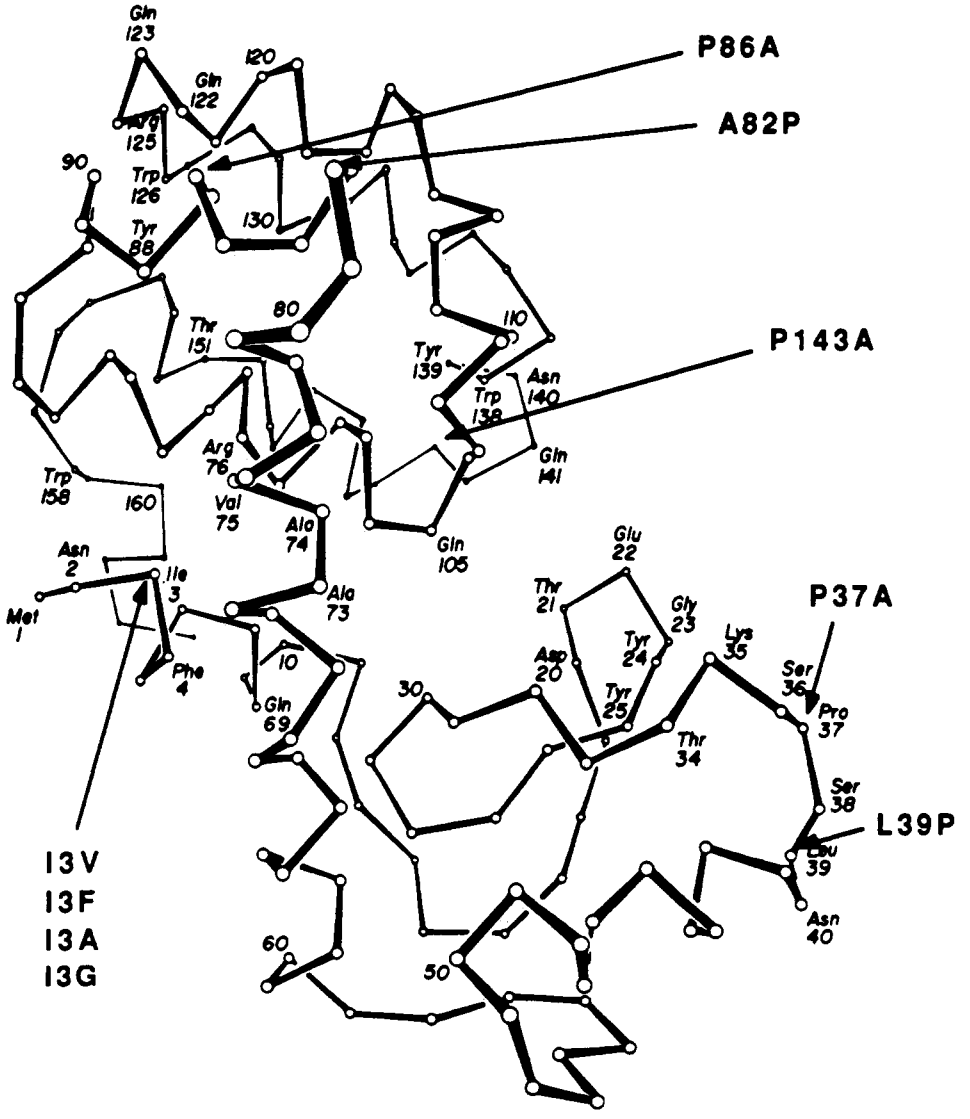

FIGURE 2: The location of the amio acid substitutions for the nine mutants of this study. Figure courtesy of B. W. Matthews and associates.

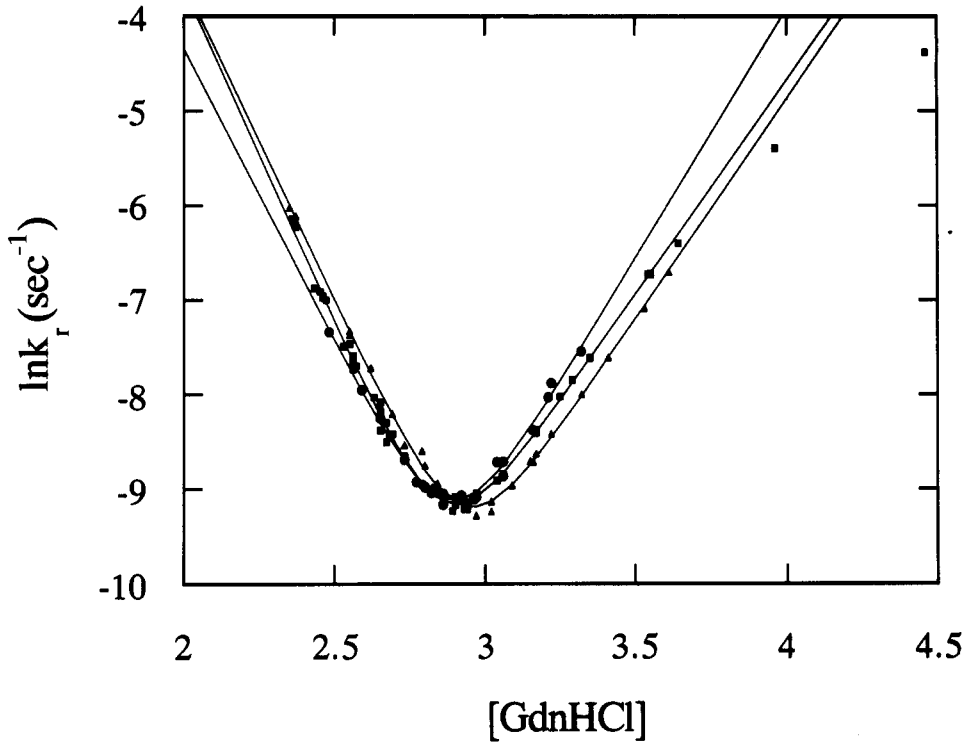

FIGURE 3: The first-order rate constant for wild-type T4 lysozyme (■) and two neutral mutants A82P (▲) and P37A (●) as a function of guanidinium chloride concentration (pH = 5.7; T = 285 K). The points are experimental; the curves are calculated from eq 4 with the parameters in Table II.

represent the fastest rates which were observed in the whole series and may be caused by the long dead times of the manual procedures. Whatever their origin, these points were not used in the chevron analysis. The mutant method depends on comparisons with the wild-type protein, and we have no matching data for the mutants in this time and concentration range.

The rate parameters obtained from the least-squares analysis are given in the top row of Table II. The free energies of activation extrapolated to zero denaturant concentration, $\Delta G^{\circ\ddagger}$, were obtained from the Eyring formula (see the preceding section for clarification). The values of $\Delta G^{\circ\ddagger}$ are subject to the same kind of extrapolation errors as equilibrium determinations, which have been thoroughly discussed (Pace, 1986; Santoro & Bolen, 1988; Schellman, 1990). We consider them to be rough estimates. On the other hand, the mean slopes of the folding and unfolding branches of the curve can be determined with good accuracy.

A second method for determining the equilibrium constant is from the forward and reverse rate constants via the relation $K = k_1/k_2$. To our knowledge, these two methods have not been compared before for this type of problem. For this reason, we give free energy values obtained by both methods in the rightmost columns of Table I. The agreement is in general very good, but the two methods are only partially independent. The t = 0 point for each kinetic run is obtained by time extrapolation with the kinetic method and by baseline extrapolation with the equilibrium method. Extrapolated baselines in equilibrium experiments can be thought of as the circular dichroism of kinetically frozen species, or of species at t = 0. The agreement is, however, a strong validation of the two-state linear model which is used. We take as best values the averages obtained from the two methods. It is these average values which appear, for example, in Figure 5.

According to the thermodynamic interpretation of transition-state theory, the limiting slopes of the folding and unfolding branches of the chevron plot provide values for the change in the free energy of solvation which occurs in going from the native state to the transition state and from the transition state to the unfolded state, respectively (Chen et al., 1989). The sum of these quantities, with a negative sign for the folding solvation free energy, must equal the equilibrium quantity $RT\Delta\beta^{\circ}$. The ratio $\Delta\beta_1^{\ddagger}/\Delta\beta^{\circ}$, symbolized by $\alpha^{\ddagger}$ is a measure of the fraction of the total free energy of solvation that is achieved in the transition state relative to the unfolded state and provides information about the solvation of the transition state relative to the longer-lived native and unfolded states (Chen et al., 1989). It is generally assumed that increased solvation free energy implies an increased area of contact with the solvent, though freakish models can be invented for which this would not be the case. Tanford had earlier proposed using the ratio of the slope of ln k_1 vs C and the slope of ln K vs C as a measure of exposure of the transition state to solvent (Tanford, 1970). Values of $\alpha^{\ddagger}$ are entered in the last column of Table II. For the wild-type protein, $\alpha^{\ddagger}$ is 0.39, distinctly larger than the value of 0.21 found for the

Table II: Kinetic Properties of the Proteins[a]

protein	$\Delta G_1^{\circ\ddagger}$	$RT\,\Delta\beta_1^{\circ\ddagger}$	$\Delta G_2^{\circ\ddagger}$	$RT\,\Delta\beta_2^{\circ\ddagger}$	τ_{max}(min)[b]	$\alpha^{\ddagger}$[c]
WT	29.5	−2.55	10.7	4.02	164	0.39
A82P	30.0	−2.63	11.1	3.81	178	0.41
P37A	30.8	−2.99	12.1	3.51	160	0.46
P86A	26.1	−1.70	7.6	4.60	25	0.27
L39P	29.0	−3.12	13.7	3.04	30	0.51
P143A	25.2	−1.60	8.9	4.82	28.5	0.25
I3V	30.0	−2.88	12.2	3.53	112	0.45
I3F	25.1	−1.55	8.0	5.09	25	0.23
I3A	25.1	−1.53	9.1	4.92	35	0.24
I3G	23.9	−1.28	9.4	5.00	18	0.20
13C-C97/C54T[d]	23.5	−0.98	10.3	3.74		0.21

[a] Conditions and units are the same as those in Table I. [b] τ_{max} is the maximum relaxation time. It is equal to $1/k_r$ at the minimum of k_r. [c] $\alpha^{\ddagger}$ is the ratio of $\Delta\beta_1^{\circ\ddagger}$ to $\Delta\beta^{\circ}$. See text. [d] Experiments done at pH 5.0.

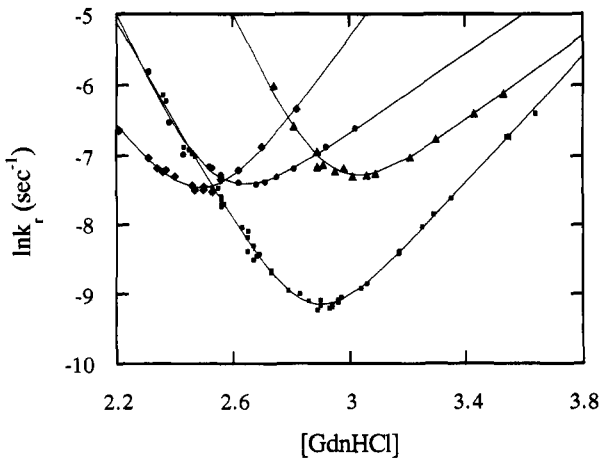

FIGURE 4: The first-order kinetics of wild-type T4 lysozyme (■) and the mutants P86A (▲), P143A (●), and L39P (◆). See legend of Figure 3.

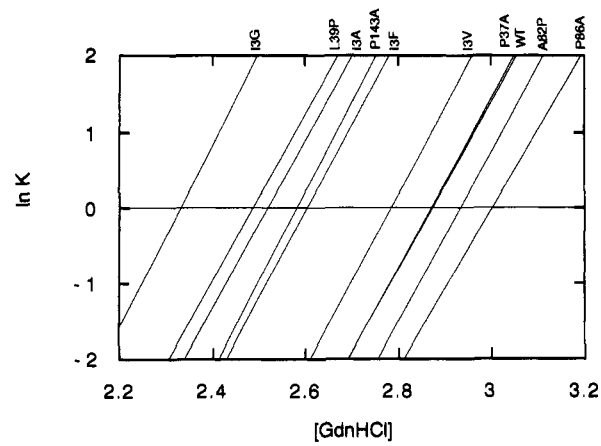

FIGURE 5: The log of the equilibrium constant vs the concentration of guanidinium chloride for the proteins in this study.

protein with the 3–97 disulfide bridge (Chen et al., 1989). The transition state as well as the unfolded state of WT has greater exposure to solvent than the S–S mutant as measured by the change in solvation free energy.

Under the most optimal conditions (T = 12 °C, pH ≅ 5–6, ionic strength ≅ 0.1 – 0.2 M), the wild-type protein has a relatively high stability ($\Delta G°$ = 18.2 kcal). It also has a very high barrier against unfolding ($\Delta G^{\ddagger}$ = 29.5 kcal), which is lowered by the addition of guanidinium ions, and a low barrier for folding which is raised by the addition of guanidinium ions. Presumably, the guanidinium interacts favorably with the partially unfolded transition state on the unfolding pathway, while in the reverse direction guanidinium ions must be removed from the unfolded state to get to the transition state. Thus, the decrease in stability in guanidinium chloride is caused by a combined increase in k_1 and a decrease in k_2.

Selection of Mutants. In addition to the wild type, we report data on two classes of mutants.

(1) Proline interchanges: These had been prepared as a part of the kinetic studies reported earlier (Chen et al., 1989). We were puzzled by the absence of an observable proline-isomerization step in the folding of I3C-C97/C54T, which displayed a single relaxation time for folding and refolding under all conditions. To explore this situation further, we investigated two mutants with known structure, P86A [J. Bell, unpublished experiments (coordinates are available on request at the Institute of Molecular Biology, University of Oregon)] and A82P (Matthews et al., 1987), and three new mutants, P37A, L39P, and P143A, which were prepared for this purpose. This gave us three mutations in which a proline had been replaced by an alanine, one in which an alanine had been replaced by a proline, and another in which a leucine had been replaced by a proline.

(2) Four replacements at site 3 which had been studied by Matsumura as part of an investigation of the effect of hydrophobicity on protein stability (Matsumura et al., 1988).

The positions of these nine amino acid replacements are shown in Figure 2 in a three-dimensional representation of T4 lysozyme using the one-letter notation to indicate the position and nature of the replacement.

Proline Interchanges. Two of the A–P interchanges, A82P and P37A, are essentially neutral mutations. This can be seen from the kinetic behavior shown in Figure 3, which is very close to that of the wild type, and from the equilibrium and kinetic entries of Tables I and II.

The results of proline interchanges at positions 39, 86, and 143 are shown in Figure 4. These mutations deviate markedly from the wild type, especially on the unfolding branch of the curves. We look first at the parameters of Tables I and II which are obtained by extrapolating to zero guanidinium concentration. These reveal the following: P86A is essentially a kinetic mutant. Both the folding and unfolding reaction are accelerated, but the equilibrium constant differs little from that of the wild type. L39P is a k_2 mutant. Its lower stability comes almost entirely from a diminished rate of folding. For P143A, both rate constants are increased (lower free energy of activation) with a net 1-kcal loss in stability. Since all of the numbers on which these conclusions are based are obtained by extrapolation, comparisons must be made at best at a semiquantitative level.

In Figure 5 we have plotted ln K for unfolding as a function of guanidinium chloride concentration for the principal mutants of this study. They all have essentially the same slope. This is reflected in the values shown for $RT\Delta\beta°$ in Table I. The total variation in these slopes is between +7% of the mean (P86A) and −7% (I3G). Since experimental error is probably of the order of 5%, we conclude that there is no evidence that the changes in slope go clearly beyond those intrinsically expected for the substitution of a single amino acid in a protein of this size. If we dismiss accidental compensations between the native and unfolded states, the conclusion is drawn that all the folded states and all the unfolded state of all these proteins are similar with regard to exposure to solvent. Since from X-ray structures and spectroscopic measurements in solution we conclude that all the native molecules are folded in essentially the same way, this means that all the unfolded molecules have distributions of conformations which are similar with regard to solvent exposure. This is what one would anticipate, since all of these proteins differ from the wild type by only a single amino acid out of 164 and the contact area of a single amino acid is very small compared to that of the entire protein.

This is not, however, a general result. Shortle and coworkers, who have studied the unfolding of a number of mutants of staphylococcal nuclease, have observed quite different slopes of $\Delta G°$ for different mutants and have interpreted this as evidence for the variability of the structure of the unfolded state for this protein (Shortle & Meeker, 1986; Shortle et al., 1989). Staphylococcal nuclease unfolds at very low concentrations of guanidinium chloride, i.e., a poor solvent for an unfolded protein, so that sequence-dependent, variable structures are more likely. Our studies reinforce Shortle's interpretation since they show that single amino acid changes in the primary structure do not lead intrinsically to observable differences in slope. Secondary and tertiary effects, caused

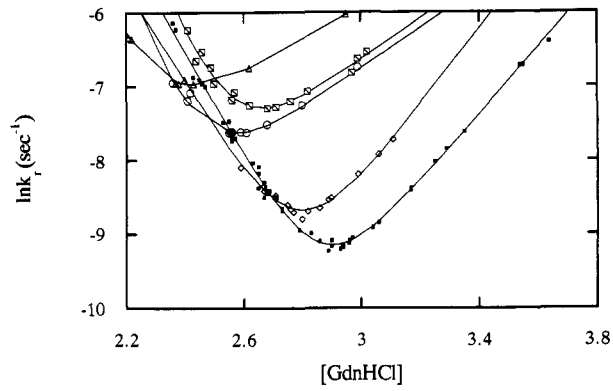

FIGURE 6: The first-order kinetics of mutations at position 3: I3V (◇), I3F (□), I3A (○), I3G (△), and the wild type (■). See legend of Figure 3.

by the mutation, must be invoked.

Changes in $\Delta\beta°$ can also be found among T4 lysozyme mutants. For example, 13C-C97/C54T with a disulfide bridge has a decreased value of $-\Delta\beta°$, a presumed indication of an unfolded state which is less exposed to solvent contacts (Chen et al., 1989).

By contrast, Figure 4 and Tables I and II indicate that point mutations produce marked changes in the slopes of ln k_r vs C plots. Since the data of Figure 5 indicate that mutations do not affect the overall solvation free energy and the data of Figure 4 indicate marked changes for the kinetics for the mutants P143A and P86A, we conclude that it is the solvation of the transition state that is sensitive to these mutations. We also learn that a casual visual examination of chevron plots can be misleading. To all appearances, the slopes in Figure 4 for P143A and P86A have changed significantly on the unfolding branch and very little on the folding branch. But since the total free energy of solvation going from the folded to unfolded form ($\Delta\beta°$) is essentially constant and $\Delta\beta° = \Delta\beta_1^{°‡} - \Delta\beta_2^{°‡}$, any decrement in the unfolding solvation must be compensated by an increment in the folding solvation and vice versa. One sees this clearly in Table II, where positive changes in $\Delta\beta_1^{°‡}$ are compensated by negative deviations in $\Delta\beta_2^{°‡}$ and vice versa. It is more difficult to see (literally) a given numerical change in slope for a steep curve than for a shallow one.

Mutations at Position 3. The second series of experiments were performed on a set of mutant proteins prepared by M. Matsumura and used by him and his co-workers in a study of the contribution of hydrophobic interactions to protein stability (Matsumura et al., 1988, 1989). Eleven amino acid substitutions were made for the isoleucine at position 3 in order to provide a thorough study of the effect of substitutions at this site on the structure and stability of the protein. Since I3 is 80% buried in a hydrophobic pocket, the results were correlated with the hydrophobicity of the side chains. Four of these proteins, with V, F, A, and G substituted for I, were generously provided by the indicated authors. Crystal structures have been published for the Y and V substitutions, but the other mutants did not crystallize. The crystal structures and the reason for the failure of the other mutants to crystallize have been presented in detail (Matsumura et al., 1989).

The results for these four proteins (I3V, I3F, I3A, and I3G) are shown in Figure 6, and the derived thermodynamic and kinetic parameters are given in Tables I and II. It is to be noted that the chevron plots for these experiments are not as fully developed as is desirable. Most of the data is crowded near the vertex. The main effect of the lack of faster data points will be greater uncertainties in the extrapolation to zero concentration, i.e., in the values of $\Delta G_1^‡$ and $\Delta G_2^‡$. These experiments were done with manual mixing techniques and gave reliable data only on a very slow time scale.

The curves of Figure 6 fall into two categories. The kinetic behavior of I3V is not too far removed from that of the wild type and has the same shape and orientation. This means that the solvation free energy for the transition state is the same as for wild type. The least-squares analysis of the curve indicates that it is a k_2 (or folding) mutant. From the previous investigations, it is already known that this mutant has a crystal structure which is very much like that of the wild type and a stability which is commensurate with expectations for removing one methylene group from the hydrophobic cavity.

The other three proteins resemble P143A. Both k_1 and k_2 are increased over the wild type, and the transition-state solvation free energy is strongly diminished. These mutants have a significant loss in stability. Though there are no crystal structures for these proteins, the A and G residues will clearly fail to fill the hydrophobic pocket in which the I3 side chain normally resides.

The results of a previous study on the mutant I3C-C97/C54T are included at the bottom of Tables I and II for the purpose of later discussion. Strict quantitative comparisons are not possible, since this protein is a double mutant, possesses a disulfide bridge, and was studied at pH = 5.0.

DISCUSSION

Measurements. Before the results are discussed and an attempt is made to interpret them, it is worthwhile to analyze the nature of the experiments and the operational meaning of the parameters we have reported in Tables I and II. Interpretation of the kinetics is dependent on the two-state model. This is implicit in all the calculations which assume that protein molecules are either folded (have one kind of CD spectrum) or unfolded (have a second type of CD spectrum). For the proteins of this paper, under the given experimental conditions, this is probably a good approximation. See Chen et al. (1989) and Chen and Schellman (1989) for a discussion of the criteria used to validate the two-state model.

The slopes of the kinetic and equilibrium constants with respect to guanidinium concentration reflect changes in solvation with changes in state. The thermodynamic description of this process for the equilibrium transition is straightforward (Schellman, 1978). The slopes depend on the number and nature of the groups exposed, but if one makes the assumption that the interior composition of proteins is approximately uniform, then they give a rough measure of the exposed surface area.

ln $k_1^°$ and ln $k_2^°$ are the intercepts of the linear limbs of the chevron plot and require care in interpretation. They are the reaction rates extrapolated to $C = 0$ but will not in general represent the rates of reaction in the absence of guanidinium chloride. If the extrapolated rate constants are taken literally, the kinetic description of this molecule is as follows. The average waiting period between unfolding events would be of the order of 300 years, followed in approximately 30 μs by refolding. *These extrapolated times are not the real rates at $C = 0$.* The folding reaction is certainly slower, and the unfolding reaction is possibly faster. These numbers demonstrate, nevertheless, that it is not reasonable to assume that the *total*, spontaneous unfolding of the molecule is a precursor of observable dynamic events such as proteolytic degradation, H exchange, and other protein reactions involving conformational changes.

Klemm et al. (1991) have performed a semiquantitative

study of the equilibrium and kinetics of urea denaturation of a number of T4 lysozyme mutants using urea-gradient gel electrophoresis. The technique permits the estimation of C_m, the cosolvation free energy, and rate constants in the neighborhood of C_m, i.e., near the apex of the chevron plot. Where there is overlap of proteins, the trends of their data and ours are in agreement. A quantitative comparison is not possible because of the difference in denaturing agents.

Further kinetic information has been obtained for the folding of T4 lysozyme by Lu and Dahlquist (1992), who have performed stopped-flow studies of the folding of T4 lysozyme mutants using fluorescence as a probe. For a cysteine-free mutant (C54T/C97A), they obtain half-times of the order of 15 ms at 38 °C and neutral pH in the absence of denaturant. Similar experiments done in the presence of urea and guanidinium chloride and extrapolated to zero denaturant concentration give time constants of the order of 10–50 ms near neutral pH at 20 °C. For folding (k_2°), our extrapolated time constants are in the microsecond range. The conclusion is that there are rate processes which are slower than our extrapolated rate k_2°, which become rate-limiting in the absence of guanidinium chloride. Similarly for k_1°, which extrapolates to centuries, there may be other processes or alternate pathways which become rate-limiting at high guanidinium concentration, and the two off-line points of Figure 3 may be an indication of this. In our studies, we are observing a process which is rate-limiting in the presence of guanidinium chloride in a time scale of minutes to hours and *the extrapolated rate constants and free energies of activation refer to the extrapolation of this process to C = 0, where it is presumably no longer rate-limiting.* This is a valid procedure for obtaining kinetic information, and its value depends on whether or not the processes we are studying lie on the kinetic pathway for folding in the absence of denaturing agents. Evidence for this comes mainly from comparison with numerous other related studies.

In contrast to k_1° and k_2°, and the associated $\Delta G_1^{\circ\ddagger}$ and $\Delta G_2^{\circ\ddagger}$, the slopes of the chevron curves do not require extrapolation and are directly determined by the experiments. In the framework of thermodynamic transition-state theory, $RT\Delta\beta_1^{\circ\ddagger}$ and $RT\Delta\beta_2^{\circ\ddagger}$ provide a measure of the free energy of solvation of the transition state relative to the native and denatured states, respectively. Even if one does not subscribe to thermodynamic transition-state theory, it is certainly a reasonable conclusion that variations in rate, induced by adding a reagent, indicate the interaction of rate-determining species with that reagent. If the rate-determining step of protein unfolding preceded any initial unfolding event, e.g., the breaking of an internal and unexposed bond or set of bonds, one would expect to see little dependence of the rate on solvent conditions.

Two Broad Kinetic Classes. If one plots any one of the kinetic or equilibrium parameters against another, the 10 proteins of this study generally fall into two classes. Two such plots are shown in Figure 7. The group labeled class I includes the wild type and mutants which show relatively small changes in the parameters from the wild-type protein (A82P, P37A, I3V, and possibly L39P); those in class II have smaller values for $\Delta G_1^{\circ\ddagger}$ and $\Delta G_2^{\circ\ddagger}$ (faster folding and unfolding), smaller exposure of the transition state to solvent, and lower stability (I3F, I3A, I3G, P143A, P86A). Crystal structures have been obtained for all of the proteins except I3F, I3A, and I3G. Where data are available, it is found that the structures for proteins in class I (except L39P) resemble those of the wild-type protein, whereas those of class II show deviations in regions of the protein close to the site of the mutation. This clear separation into two groups will probably disappear as more mutations are accumulated, but the results indicate that the lowering of the stability of the protein, increased rates of folding and unfolding, and reduced exposure of the transition state to solvent are all correlated. Neglecting changes in experimental conditions, the previously studied mutant I3C-C97/C54T falls into class II as far as its kinetic and equilibrium behavior is concerned. No structure is as yet available.

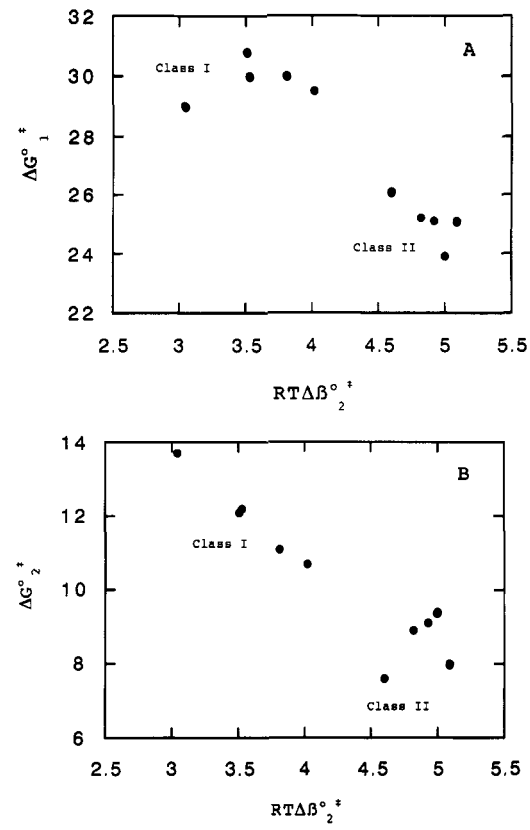

FIGURE 7: Examples of the correlations among the kinetic and equilibrium parameters for all of the proteins: (A) $\Delta G_1^{\circ\ddagger}$ vs $RT\Delta\beta_2^{\circ\ddagger}$,; (B) $\Delta G_2^{\circ\ddagger}$ vs $RT\Delta\beta_2^{\circ\ddagger}$. The meanings of the symbols are given in the abbreviations footnote.

A Kinetic Mutant. One would like to match the kinetic and thermodynamic data on the mutants with effective structural models as a first stage in unscrambling the factors which affect folding and unfolding. These models could then be tested by designing further mutants. This is not possible for many of the proteins in this study. We discuss the case of P86A first because a simple and reasonable model is suggested by the data.

Proline 86 occurs in the middle of the 82–90 helix in wild-type T4 lysozyme, and Alber et al. (1988) introduced a series of 10 amino acid replacements with varying helix-forming tendencies at this position to see if a more stable protein could be produced. These substitutions resulted in small decreases in stability as measured by changes in transition temperature. In the same paper, crystal structures of six of these mutants are reported. Subsequently, the crystal structure of P86A has been obtained [J. Bell, unpublished experiments (coordinates are available on request at the Institute of Molecular Biology, University of Oregon)]. These structures indicate that the substitutions result in the removal of helical distortions caused by the proline but that the straighter, less deformed helix has less favorable tertiary interactions. These effects essentially cancel one another. Stated crudely, the proline-deformed helix fits into the tertiary structure of the rest of the protein better than the more ideal helix obtained by removal of the proline.

Since the P86 mutants have only very small changes in stability, they will be kinetic mutants if they have altered kinetics, and this study shows that the unfolding and refolding of P86A is significantly faster than that of wild type. This result can be understood with two reasonable postulates: (1) the 82–90 helix exists (at least partially) in the transition state; (2) a fixed tertiary structure around this helix is nonexistent or is not fully developed in the transition state. Under these circumstances, the improvements in secondary structure caused by removal of the proline would not be counterbalanced by poorer tertiary interactions. This would lower the free energy of the transition state (as was originally anticipated for the native state), and the reaction rates would be enhanced equally in both directions. It may be possible to confirm this hypothetical mechanism by the kinetic study of other mutations in the same helix.

Kinetically Neutral Mutations, A82P and P37A. When compared to the wild-type protein, these two mutants show minimal variations in structure, thermal stability, kinetic parameters, and enzymatic activity. (The activity of P37A was not measured, but it showed wild-type halos.) The free energy of stabilization and extrapolated free energies of activation are clustered near to wild-type values as are all the β parameters. Both involve an A $\leftrightarrow$ P interchange of a residue which is on the surface of the protein and is exposed to solvent. Neither residue has a low temperature factor so that the flexibility criterion for stable mutational substitutions is satisfied (Alber et al., 1987). Residue 82 in the wild type is in the first α-position at the N-terminal end of an α-helix. This is a position where proline residues are frequently found in proteins. In addition, the (ϕ,ψ) values for A82 are (−61,−28) which is a low-energy region for proline residues. Residue 37 is in an irregular region of the protein, completely exposed to solvent and with (ϕ,ψ) values compatible with either A or P residues. Thus, these two mutants meet the requirements which have been established for producing minimal changes in stability (Alber et al., 1987), and the present investigation shows that the kinetic effects of these mutations are also small.

P143A. This is a proline–alanine interchange at the first α-helical residue on the N-terminus of a helix. The results, however, are quite different from A82P, which also occurs at this helical position. Relative to the wild type, P143A shows a decrease in stability, marked changes in both unfolding and refolding kinetics, decreased exposure of the transition state to solvent, and significant structural displacements as determined by X-ray diffraction. Even though residue 143 is at the end of a helix and is close to the surface of the protein, it has low thermal coefficients for its atoms and is held relatively rigidly in place compared to the neutral substitutions. There is a movement of about 0.6 Å in the backbone, a shift of the 143–155 helix, and a new solvent peak, indicating changes in both structure and solvation. Details will appear elsewhere [H. Nicholson, unpublished experiments (coordinates are available on request)]. These distortions raise the free energy of the native state relative to the unfolded state by about 1.5 kcal. This in itself would account for a decreased barrier for unfolding (as is observed), if one could assume that the transition state was unchanged. This, however, is not the case. P143A is one of the class II proteins which shows not only more rapid unfolding rates but also large changes in the relative solvation of the transition state, indicating significant mechanistic changes.

L32P. Like A82P, this mutant is also the substitution of a proline at the first α position at the N-terminus of a helix. In general, the replacement of leucine residues in proteins causes considerable changes in stability and structure. This is because the large, hydrophobic side chain of this amino acid is normally buried in the interior of proteins, and the substitution of a smaller or more polar residue leaves cavities or repulsive interactions. L39 is only partially buried, but it shows the same structural effect. There is a contact between L39 and Y25 which is lost in the replacement. The structure [H. Nicholson, unpublished experiments (coordinates are available on request)] shows that the substituted proline residue attempts to restore this favorable contact by moving closer to Y25, but this leads to unfavorable distortions in the helix.

L39P is the only one of the 10 proteins which does not fit too well into the classes defined above, which are at any rate only a temporary way of organizing the data. Its transition state is even more exposed than that of the wild type, and it is the only protein which has a slower unfolding rate combined with a marked lowering of stability. More information will have to be obtained before we can attempt to understand such differences on a structural basis.

I3V. This and 12 other replacements at position 3 have been reported by Matsumura et al. (1988, 1989). Crystal structures are available for the wild type, I3V, and I3Y. In the wild type, the isoleucine side chain is buried in a hydrophobic pocket in contact with M6, L7, I100, and the main chain of C97. Assuming that the local structure is not changed when I is changed to V, A, and G, Matsumura et al. have compared the surface area buried in the native fold for the subset of proteins I, V, A, and G, and the same subset was selected for our kinetic measurements. The exposed surface area of these residues was calculated for the fully extended state and for the native state of the protein where the side chains are located in the hydrophobic pocket (Matsumura et al., 1989). The numbers obtained were 131, 111, 72, and 40 Å^2, respectively for I, V, A, and G. It was found that the decrease in the stabilization free energy of the three mutants is directly proportional to the loss of hydrophobic contact area with a proportionality constant very close to that estimated by Chothia from transfer free energies (Chothia, 1974). The relatively small decrease in hydrophobic contact of the I $\rightarrow$ V exchange correlates with the relatively small decrease in stability of this mutant. This investigation shows that the changes in kinetic parameters for folding and unfolding are also comparatively small. From the table, it appears that I3V is a k_2 mutant. The extrapolated rate of unfolding and the exposure of the transition state to solvent are about the same as those of the wild type, but the extrapolated folding rate is somewhat slower.

I3A, I3G, and I3F. These three mutants all fall in class II. They are less stable than the wild type; they have markedly reduced exposure of the transition state to solvent and markedly faster unfolding rates; their extrapolated folding rates are also faster, but it is to be noted that the folding branches of the curves for I3A and I3G cross the curves for the wild type. The experimental points for these curves are more restricted than those discussed earlier, so the extrapolated rates and $\Delta G^{\circ\ddagger}$ values are less dependable. The A and G replacements have been mentioned in the previous section. There are no crystal structures for these two mutants, but the stability of the series I, V, A, G is in the order of the buried surface area of the side chains, if one assumes the native structure for the remainder of the molecule (Matsumura et al., 1988). The phenylalanine structure was not compared in this way because the crystal structure of I3Y has been determined and it is found that the tyrosine aromatic ring has pivoted out of the hydrophobic pocket to make contact with a neighboring protein molecule in the crystal.

Though the I → V exchange produces relatively small differences in the structure, stability, and kinetic properties of T4 lysozyme, presumably because of the small changes in shape, volume, and hydrophobicity, the introduction of F, A, and G generates proteins which deviate significantly from the wild-type protein (see Tables I and II). If we take the unfolded molecule as the standard state for comparison, the free energy of the native form of these mutants is raised because of less extensive (or effective) burial of the hydrophobic groups and because of the introduction of small strains, pockets, and rearrangements in the native state. On the other hand, the transition state is stabilized relative to the unfolded state, leading to a lower barrier. The combination of a raised native state and a lowered transition state results in a markedly enhanced rate of unfolding. The lowered free energy of the transition state of these mutants, as well as that of P86A, may be the result of the reduced exposure to solvent (see the last column of Table II).

What Process Is Being Studied? The question which invariably follows an experimental study of rates is the nature of the rate-determining step. In the present case there are two questions: the nature of the rate-determining step of the process we have studied, i.e., that which occurs in the presence of a denaturant like guanidine, and the relation of this process to the folding under more normal or in vivo conditions. It is clear from the nature of the chevron plot that the effect of the denaturant is to slow the folding and accelerate the unfolding reactions until above C_m the folded protein is unstable. We can study the relative effects of denaturants on the native and unfolded states, and by using transition state theory we obtain information about the effect of denaturants on the transition state. But there are presumably many transient states on the folding pathway and these may be affected differently by the presence of a denaturant. In particular, the transition state or the rate-determining step may differ in the presence of added destabilizing reagents such as urea or guanidine salts.

In fact, we have evidence that this is the case in the experiments of Lu and Dahlquist (1992), which show that folding rates in the absence of denaturants are in the neighborhood of 10–50 ms at temperatures of 32 and 38 °C. This is in agreement with a number of other studies of the folding of small proteins by rapid techniques. Characteristic times for the *fastest* steps tend to be in the 10–100-ms region. Our measurements, on the other hand, extrapolate to the 10-μs range on the basis of the linear model. Since we are extrapolating to find the logarithm of k_2°, errors will be large in k_2° itself. The error analysis discussed in the section on data interpretation indicates that our extrapolations of rate constants should not be off by more than an order of magnitude for the three worst cases and should be considerably better than this for the other proteins (factors of 1.5 to less than 5). Since the deviation from the fluorescence extrapolations is 2–3 orders of magnitude, we conclude that the process we are studying extrapolates to a step which is not the rate-limiting process in the absence of denaturant. This type of conclusion is not new. C. R. Matthews' group has observed a number of instances where ln k levels off to a constant value at low concentrations of denaturant rather than continuing the chevron pattern as a linear function of concentration. This means that another process becomes rate-limiting which is independent of denaturant concentration. In the case of tryptophan synthetase, it has been concluded that the denaturant-independent step is the formation of an intermediate and that the denaturant-dependent step is the conversion of this intermediate to the native protein (Matthews et al., 1983; Beasty et al., 1986; Hurle et al., 1987; Matthews & Hurle, 1987).

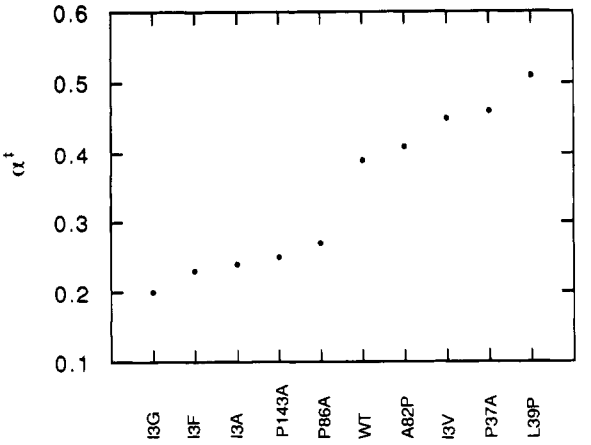

FIGURE 8: The solvent exposure parameter $\alpha^* = \beta_1^{\circ *}/\beta^\circ$. See explanation in text.

The kinetics of folding of T4 lysozyme are simpler than those for tryptophan synthetase, since the reactions we observed are two-state and we have no evidence for partially stable intermediates. There is, however, no lack of suggestions for slower steps on the pathway. These include intramolecular long-range diffusion involving the long relaxation times of a polymer chain, the diffusion–collision model (Karplus & Weaver, 1976), and the sorting out of native structure from a compact, partially ordered structure [the framework model (Kim & Baldwin, 1982) and molten globule model (Kuwajima, 1989; Ptitsyn, 1987; Ptitsyn et al., 1990)].

Further information on the guanidine-dependent step which dominates the measurements of this paper may be obtained by considering the effect of the mutations on the interaction parameters $\Delta\beta_1^{\circ *}$ and $\Delta\beta_2^{\circ *}$. As outlined in the Results section, there is a considerable variation in the α^* factor (the ratio of $\Delta\beta_1^{\circ *}$ to $\Delta\beta^\circ$) for the various mutants. This variation is depicted in Figure 8. For convenience, we earlier divided this behavior into two classes, but it can be seen that α^* varies almost continuously from 0.20 to 0.50. This is an amazingly large range. It indicates a variation in the solvent-contact free energy of different forms of the transition state, which is 30% of the total change in contact free energy of the unfolding reaction. This is in complete contrast to the equilibrium properties, where changing a single amino acid makes only a small change in $\Delta\beta^\circ$, which is commensurate with the very small change in the total surface of a protein caused by the change of one amino acid.

What could this result mean? It could mean that there are different transition states for each protein indicating different folding pathways. An alternative possibility is that the proteins follow the same general pathway but that the point of highest free energy (the transition state) is reached at different stages for different proteins. Resolution between these two concepts cannot be made with the data at hand, but the systematic variation of the position of mutations of the protein may yield further information. It at any rate seems clear that the folding process studied in guanidinium chloride solutions is associated with the formation of a compact form of the protein with a reduced area of contact with solvent and that the structure of the compact form (as measured by the solvent-contact parameter) varies very considerably with different point mutations.

Can These Data Be Fit into Current Models for Protein Folding? The main qualitative conclusions of this study are (1) The rate-determining process of protein folding in guanidinium solutions extrapolates to a process that is fast com-

pared with the folding reaction in the absence of guanidinium chloride; (2) The process studied in guanidine is associated with the formation of solvent-shielded and presumably compact structures of the protein; (3) The structure of the compact intermediate, as measured by its solvent-contact parameter, is remarkably sensitive to point mutations; (4) The changes in kinetic properties can usually be correlated with structural changes in the protein, though how these correlations fit in with details of the kinetic pathway or the structure of intermediates still remains unclear for most of the mutants.

There are a number of protein-folding models which postulate the early formation of compact, partially solvent-shielded structures in the folding pathway, and the present results can probably be brought into accord with any of them. The relation of these models have been the subject of a very clear review by Baldwin (1989). For discussion, we select the framework model (Kim & Baldwin, 1982) and/or the molten globule model (Kuwajima, 1989; Ptitsyn, 1987; Ptitsyn et al., 1990). The distinction between these models appears to be becoming blurred in recent studies and apparently depends on whether one considers secondary interactions to precede hydrophobic compaction, or vice versa. Dill and co-workers have concluded that the formation of secondary structure is a necessary consequence of compaction of the protein chain (Chan & Dill, 1989). The diffusion–collision model (Karplus & Weaver, 1976) and the framework model (Kim & Baldwin, 1982) require the subsequent stabilization of secondary structures by tertiary interaction. The distinction is of importance only if the two pathways lead to different structures, and it will not affect the following discussion.

We now cite relevant studies from the literature. (1) The molten globule form has been detected as a stable state of a number of proteins under certain experimental conditions. It is identified experimentally as a compact state with a partial complement of secondary structure but no evidence of a fixed tertiary structure. (2) It has also been identified as a transient state in the folding of carbonic anhydrase B (Dolgikh et al., 1984), hen egg-white lysozyme (Ikeguchi et al., 1986), and ferricytochrome *c* and β-lactoglobulin (Kuwajima et al., 1987). Its formation is followed by a slower transformation to the native protein. (3) An NMR trapping technique has been developed which provides a kinetic curve of the availability of protons for exchange during the folding process. The technique may be applied to all individual amide protons, provided their exchange is slow enough in the native form. This technique has now been applied to ribonuclease A (Udgaonkar & Baldwin, 1988), horse heart cytochrome *c* (Roder et al., 1988), barnase (Bycroft et al., 1990), lysozyme (Miranker et al., 1991), and T4 lysozyme (Lu & Dahlquist, 1992). All these cases reveal the presence of a transient state with a secondary structure that is a subset of the secondary structure of the native protein. (4) NMR studies of the molten globule state of α-lactalbumin (Baum et al., 1989) and apomyoglobin (Hughson et al., 1990) demonstrate a pattern of partial secondary structure similar in kind to the transient intermediary states just cited. (5) Shortle and co-workers have studied the equilibrium between the native and "unfolded state" of staphylococcal nuclease as a function of guanidine (Shortle, 1986; Shortle & Meeker, 1986; Shortle et al., 1989). In contradistinction to our results (Figure 5), the slope of the curves of ln K vs C was found to vary remarkably with different point mutations. It was concluded that the unfolded state of this protein is compact and maintains a number of intramolecular interactions which vary sensitively with single changes in sequence. This result should be compared with Figure 8.

We can now try to fit our results into a standard model which applies to small proteins like ribonuclease A, lysozyme, or T4 lysozyme, which appear to be two-state reactions once proline isomerizations have been eliminated or taken into consideration. If we consider that there is an intermediate, compact state as in the framework or molten globule models, the scheme can be written as (Baldwin, 1989)

$$U \underset{\text{coalescence}}{\leftrightarrow} A \underset{\text{consolidation}}{\leftrightarrow} N$$

where A is the compact intermediate. According to the framework or molten globule models, the U → A step involves the formation of secondary structural units held together by hydrophobic interactions. A is stabilized by hydrophobic interactions but has no hydrophobic core. In the A → N step, the secondary structure of the native form is completed, correct tertiary conformations are occupied by the side chains, and water molecules are removed to the exterior of the protein. Evidence for the existence and properties of the A form has accumulated from molten globule studies on α-lactalbumin and other proteins [see Kuwajima (1989) for a recent review] from kinetic experiments which measure the formation of secondary and tertiary structure independently (Dolgikh et al., 1984; Ikeguchi et al., 1986; Kuwajima et al., 1985) and from the NMR investigations cited above. With some systems and experimental conditions, the A form can be isolated and observed; with others, it appears as a kinetic intermediate; with still others, it has only been inferred from NMR trapping and similar experiments.

Under normal experimental conditions, it is the second step that appears to be rate-determining, so that the possible diffusion control of the U → A reaction is not observed. The A-form has an enhanced contact with the solvent relative to N and is destabilized in urea or guanidine solutions [evidence reviewed by Kuwajima (1989)]. The kinetic data presented in this paper can be brought into concordance with this model if it is assumed that the A-form becomes so unstable in guanidinium chloride solutions that its formation becomes the rate-determining step in folding. This would account for the marked dependence of the rate on denaturant concentration and the values of α^* which indicate a state characterized by a contact with the solvent that is intermediate between the native and unfolded states. A-forms have already been proposed as intermediates in the refolding of proteins in the presence of denaturants [Dolgikh et al., 1981; Kuwajima, 1989]. To date, a stable A-form has not been directly observed for T4 lysozyme. The CD spectra of both its high-temperature unfolded form as well as the guanidinium-unfolded form reveal no evidence of secondary structure. The probability of observing it will be greatly enhanced at low concentrations of denaturants where this model would predict the U → A step to be faster than the A → N step. This is essentially the procedure which was used with success by Kuwajima and co-workers for hen egg-white lysozyme.

The studies of Shortle and co-workers (Shortle & Meeker, 1986; Shortle et al., 1989) on the unfolding of staphylococcal nuclease in the presence of guanidinium chloride make an interesting comparison with our studies of the transition state in the presence of this reagent. They found that the solvent exposure parameter, $RT\Delta\beta°$, varied markedly for a number of single amino acid replacements. It was also shown that the "unfolded protein" was very compact and possessed some secondary structure (Shortle & Meeker, 1986, 1989). It was also demonstrated that this compact form did not require the presence of guanidinium chloride. Staphylococcal nuclease is an unusual protein in that its C_m for unfolding by guani-

dinium chloride is less than 1 M. These studies demonstrated that the result of unfolding this protein in low concentrations of denaturant is a compact form with the properties postulated for the A-form in the works of Ptitsyn, Kuwajima, and others. In connection with our work, a main point of interest is their finding that single amino acid replacements produce very large changes in the solvation parameter for the compact but unfolded form. Evidently, this A-form is very labile and a single side-chain substitution can produce changes in contact with solvent which are an order of magnitude larger than is possible by considering the solvent contact of the side chain by itself. This is, however, exactly the type of behavior that was found for the transition state in our studies (Figure 8) and lends support to the assumption that the transition state in guanidinium chloride resembles the A-form postulated in other studies.

As a final conclusion we see that, in the context of the framework or molten globule models, the presence of denaturants perturbs the kinetics of folding of T4 lysozyme by switching control of the rate to an earlier step which is evidently a compaction to the A-form intermediate of the framework and molten globule theories. This means that we are not studying the rate-controlling processes which are important in the absence of denaturants and in vivo. On the other hand, we are gaining information on the structure and stability of a major intermediate which displays a strong sensitivity to mutational changes and which is an essential ingredient in the overall mechanism of folding. This conclusion depends on the appropriateness of the U ↔ A ↔ N model for our system and on the accuracy of the extrapolations, which is analyzed in the section on the interpretation of chevron plots.

Acknowledgments

We thank Robert Baldwin, David Shortle, Andy Morton, and Charlotte Schellman for careful reading and helpful comments on the manuscript; Robert Baldwin for continued and needed advice and information during its preparation; Jirong Lu, Andy Morton, and Frederick Dahlquist for kinetic data prior to publication; Lawrence McIntosh for a steady input of ideas, protein mutations, and unpublished data; Brian Matthews for regular discussions and structural interpretations during the course of the work; Masazumi Matsamura for the group of protein mutants at position 3; Joan Wozniak for protein preparations and purifications; Jeff Bell for unpublished structural data on P86A; and Jeff Bell, Ray Jacobsen, and Ron Albright for assistance with the X-ray crystallography.

References

Alber, T., Bell, J. A., Dao, P. S., Nicholson, H., Wozniak, J. A., Cook, S., & Matthews, B. W. (1988) *Science 239*, 631–635.

Alber, T., & Matthews, B. W. (1987) *Methods Enzymol. 154*, 511–533.

Alber, T., Sun, D. P., Nye, J. A., Muchmore, D. C., & Matthews, B. W. (1987) *Biochemistry 26*, 3754–3758.

Anderson, W. D., Fink, A. L., Perry, L. J., & Wetzel, R. (1990) *Biochemistry 29*, 3331–3337.

Baldwin, R. L. (1989) *Trends Biochem. Sci. (Pers. Ed.) 14*, 291–294.

Baum, J., Dobson, C. M., Evans, P. A., & Hanley, C. (1989) *Biochemistry 28*, 7–13.

Beasty, A. M., Hurle, M. R., Manz, J. T., Stackhouse, T., Onuffer, J. J., & Matthews, C. R. (1986) *Biochemistry 25*, 2965–2974.

Berne, B. J., Borkovec, M., & Straub, J. E. (1988) *J. Phys. Chem. 92*, 3711–3725.

Bevington, P. R. (1969) *Data Reduction and Error Analysis for the Physical Sciences* McGraw-Hill, New York.

Bolen, D. W., & Santoro, M. (1988) *Biochemistry 27*, 8069–8074.

Bycroft, M., Matouschek, A., Kellis, J. T., Serrano, L., & Fersht, A. R. (1990) *Nature 346*, 488–490.

Chan, H. S., & Dill, K. A. (1989) *J. Chem. Phys. 90*, 492–509.

Chen, B. L., Baase, W. A., & Schellman, J. A. (1989) *Biochemistry 28*, 691–699.

Chen, B. L., & Schellman, J. A. (1989) *Biochemistry 28*, 685–691.

Chothia, C. (1974) *Nature 248*, 338–339.

Desmadril, M., & Yon, J. M. (1984) *Biochemistry 23*, 11–19.

Dolgikh, D. A., Gilmanshin, R. I., Brazhnikov, E. V., Byochlova, V. E., Semisotnov, G. V., Venyaminov, S. Y., & Ptitsyn, O. B. (1981) *FEBS Lett. 136*, 311–315.

Dolgikh, D. A., Kolomiets, A. P., Bolotina, I. A., & Ptitsyn, O. B. (1984) *FEBS Lett. 165*, 88–92.

Elwell, M., & Schellman, J. A. (1975) *Biochim. Biophys. Acta 494*, 367–383.

Hughson, F. M., Wright, P. E., & Baldwin, R. L. (1990) *Science 249*, 1544–1548.

Hurle, M. R., Michelotti, G. A., Crisanti, M. M., & Matthews, C. R. (1987) *Proteins: Struct., Funct., Genet. 2*, 54–63.

Ikeguchi, M., Kuwajima, K., Mitani, M., & Sugai, S. (1986) *Biochemistry 25*, 6965–6972.

Karplus, M., & Weaver, D. (1976) *Nature 260*, 404–406.

Kim, P. S., & Baldwin, R. L. (1982) *Annu. Rev. Biochem. 51*, 459–489.

Klemm, J. D., Wozniak, J. A., Alber, T., & Goldenberg, D. P. (1991) *Biochemistry 30*, 589–594.

Knott, G. D. (1979) *Comput. Programs Biomed. 10*, 271–280.

Kuwajima, K. (1989) *Proteins: Struct. Funct., Genet. 6*, 87–103.

Kuwajima, K., Hiraoka, Y., Ikeguchi, M., & Sugai, S. (1985) *Biochemistry 24*, 874–881.

Kuwajima, K., Mitani, M., & Sugai, S. (1989) *J. Mol. Biol. 206*, 547–561.

Kuwajima, K., Yamaya, H., Miwa, S., Sugai, S., & Nagamura, T. (1987) *FEBS Lett. 221*, 115–118.

Lu, J. R., & Dahlquist, F. W. (1992) *Biochemistry* (submitted for publication).

Matsumura, M., Becktel, W. J., & Matthews, B. W. (1988) *Nature 334*, 406–10.

Matsumura, M., Wozniak, J. A., Sun, D. P., & Matthews, B. W. (1989) *J. Biol. Chem. 264*, 16059–66.

Matthews, B. W., Nicholson, H., & Becktel, W. J. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 6663–6667.

Matthews, C. R., Crisanti, M. M., Manz, J. T., & Gepner, G. L. (1983) *Biochemistry 22*, 1445–1452.

Matthews, C. R., & Hurle, M. R. (1987) *BioEssays 6*, 254–7.

Miranker, A., Radford, S. E., Karplus, M., & Dobson, C. M. (1991) *Nature 349*, 633–636.

Muchmore, D. C., McIntosh, L. P., Russell, C. B., Anderson, D. E., & Dahlquist, F. W. (1989) *Methods Enzymol. 177*, 44–73.

Nozaki, Y. (1972) *Methods Enzymol. 26*, 43–50.

Pace, C. N. (1975) *CRC Crit. Rev. Biochem. 3*, 1–43.

Pace, C. N. (1986) *Methods Enzymol. 131*, 266–280.

Pace, C. N., & Tanford, C. (1968) *Biochemistry 7*, 198-207.
Perry, K. M., Onuffer, J. J., Gittelman, M. S., Barmat, L., & Matthews, C. R. (1989) *Biochemistry 28*, 7961-8.
Privalov, P. L. (1990) *Crit. Rev. Biochem. Mol. Biol. 25*, 281-305.
Ptitsyn, O. B. (1987) *J. Protein Chem. 6*, 273-293.
Ptitsyn, O. B., Pain, R. H., Semisotnov, G. V., Zerovnik, E., & Razgulyaev, O. I. (1990) *FEBS Lett. 262*, 20-24.
Roder, H., Elove, G. A., & Englander, S. W. (1988) *Nature 335*, 700-704.
Santoro, M., & Bolen, D. W. (1988) *Biochemistry 27*, 8063-8068.
Schellman, J. A. (1978) *Biopolymers 17*, 1305-22.
Schellman, J. A. (1990) *Biopolymers 29*, 215-224.
Shortle, D. (1986) *J. Cell. Biochem. 30*, 281-9.
Shortle, D., & Meeker, A. K. (1986) *Proteins: Struct., Funct., Genet. 1*, 81-9.
Shortle, D., & Meeker, A. K. (1989) *Biochemistry 28*, 936-44.
Shortle, D., Meeker, A. K., & Gerring, S. L. (1989) *Arch. Biochem. Biophys. 272*, 103-13.
Tanford, C. (1970) *Adv. Protein Chem. 24*, 1-95.
Udgaonkar, J. B., & Baldwin, R. L. (1988) *Nature 335*, 694-9.
Zoller, M. J., & Smith, M. (1984) *DNA 3*, 479-488.

Stabilities of Disulfide Bond Intermediates in the Folding of Apamin†

Beatrice M. P. Huyghues-Despointes‡ and Jeffrey W. Nelson*
Department of Biochemistry, Louisiana State University, Baton Rouge, Louisiana 70803-1806
Received June 21, 1991; Revised Manuscript Received October 23, 1991

ABSTRACT: Apamin is an 18-residue bee venom peptide with the sequence CNCKAPETALCARRCQQH-amide and contains 2 disulfide bonds connecting C-1 to C-11 and C-3 to C-15. In the folding of reduced, unfolded apamin to native apamin with two disulfide bonds, the one-disulfide folding intermediate states are not populated to significant levels. To study the properties of the one-disulfide intermediates, we have synthesized two peptide models to mimic the one-disulfide intermediates, Apa-1 and Apa-2, in which two cysteines in the sequence have been replaced by alanines. These peptides can form only one of the native disulfide bonds, C-1 to C-11 in the case of Apa-1 and C-3 to C-15 in the case of Apa-2. The stabilities of these disulfide bonds have been measured as a function of pH, concentration of urea, and temperature, in order to understand which contributions stabilize the disulfide-bonded structures. Using oxidized and reduced glutathione, the equilibrium constants for forming the disulfide bonds at 25 °C and pH 7.0 are 0.018 M for Apa-1 and 0.033 M for Apa-2 and show little dependence on pH or temperature. Both disulfide bonds are destabilized slightly (by approximately a factor of 2) between 0 and 8 M urea. Circular dichroism spectra indicate that although both Apa-1 and Apa-2 exhibit some structure, Apa-2 exhibits more than Apa-1. The results suggest that in the folding of apamin, the one-disulfide intermediate containing the C-3 to C-15 disulfide bond, as in Apa-2, is favored slightly. Secondary structure provides modest stabilization to this intermediate.

Understanding the folding pathway of a protein involves identifying and characterizing the intermediates found along the pathway. This goal is made difficult because of the cooperativity of protein folding: any intermediate state will only be populated transiently, with only fully folded and denatured states significantly populated at equilibrium. Despite these difficulties, the existence of folding intermediates has long been known: kinetics measured by a single probe often exhibit multiple phases, and monitoring different probes often yields different kinetic curves (Kim & Baldwin, 1980).

Several techniques have been developed to selectively populate transient intermediates. Incorrect proline isomers can trap a nativelike intermediate along the slow folding pathway of pancreatic ribonuclease A (RNase A)[1] (Cook et al., 1979), and the stability of such an intermediate to GdmCl denaturation can be measured by double-jump experiments starting from unfolded protein (Grafl et al., 1986). A folding intermediate along the fast folding pathway of RNase A can also be transiently populated and examined by double-jump experiments starting with native protein (Hagerman et al., 1979). The nature of the structure formed along the folding pathway can be inferred by trapping exchangeable amide protons along the peptide backbone and examining by NMR when each amide proton is protected from exchange. Pioneering studies have been carried out for RNase A (Udgaonkar & Baldwin, 1988) and cytochrome *c* (Roder et al., 1988). Excellent reviews of protein folding have recently been published (King, 1989; Kim & Baldwin, 1990).

† Financial support was provided by NIH Grant GM 39615. Acknowledgment is made to the donors of the Petroleum Research Fund, administered by the American Chemical Society, for support of this work.
* To whom correspondence should be addressed.
‡ Present address: Department of Biochemistry, Stanford University School of Medicine, Stanford, CA 94305.

[1] Abbreviations: GSH, reduced glutathione; GSSG, oxidized glutathione; Ellman's reagent or DTNB, 5,5′-dithiobis(2-nitrobenzoic acid); EDTA, ethylenediaminetetraacetic acid; DTT, dithiothreitol; TFA, trifluoroacetic acid; TEA, triethylamine; GdmCl, guanidinium chloride; HPLC, high-performance liquid chromatography; CD, circular dichroism; [θ], mean residue ellipticity; NMR, nuclear magnetic resonance; RNase A, pancreatic ribonuclease A; BPTI, bovine pancreatic trypsin inhibitor.

0006-2960/92/0431-1476$03.00/0