Protein Denaturation: A Small-Angle X-ray Scattering Study of the Ensemble of Unfolded States of Cytochrome *c*†

Daniel J. Segel,‡ Anthony L. Fink,§ Keith O. Hodgson,|| and Sebastian Doniach*,‡

Departments of Physics and Chemistry, Stanford University, Stanford, California, 94305, Department of Chemistry and Biochemistry, University of California, Santa Cruz, California, 95064, and Stanford Synchrotron Radiation Laboratory, Stanford, California 94309

Received March 9, 1998; Revised Manuscript Received June 16, 1998

ABSTRACT: Solution X-ray scattering was used to study the equilibrium unfolding of cytochrome *c* as a function of guanidine hydrochloride concentration at neutral pH. The radius of gyration (R_g) shows a cooperative transition with increasing denaturant with a similar C_m to that observed with circular dichroism. However, the lack of an isoscattering point in the X-ray scattering patterns suggests the equilibrium unfolding is not simply a two-state process. Singular value decomposition (SVD) analysis was applied to the scattering patterns to determine the number of distinct scattering species. SVD analysis reveals the existence of three components, suggesting that at least three equilibrium states of the protein exist. A model was employed to determine the thermodynamic parameters and the scattering profiles of the three equilibrium states. These scattering profiles show that one state is native (N). The other two states (U_1, U_2) are unfolded, with U_2 being fully unfolded and U_1 having some residual structure. Using the thermodynamic parameters to calculate fractional populations, U_1 is maximally populated at intermediate denaturant concentrations while U_2 is maximally populated at high denaturant concentrations. It is likely that there is a multiplicity of denatured states with U_1 and U_2 representing an average of the denatured states populated at intermediate and high denaturant concentrations, respectively.

Many small globular proteins in solution fold reversibly between equilibrium states when the solvent conditions are changed from high to low denaturant concentration (*1*). To understand the kinetics of protein folding, it is important to find ways of characterizing and quantifying the kinds of conformations which are populated during the folding process. NMR studies provide good evidence that the protein conformations populated under renaturing conditions (i.e., low or 0 denaturant concentration) are very close to those measured by X-ray crystallography (*2*−*4*). However, there is relatively little information on the nature of the various conformations which are populated in the denatured state (*5*, *6*).

Equilibrium studies of protein denaturation using local structural probes such as UV circular dichroism and fluorescence lead to titration curves of probe signal vs denaturant concentration which can often be fit by a two-state model. In this model it is assumed that the equilibrium state of the protein at given denaturant concentration may be represented as a sum of two distinct populations of protein conformations: one subensemble is centered around the native state while the second is an ensemble of unfolded conformations. In such two-state fits, each subensemble is allowed a linear "baseline" dependence of the probe signal as a function of denaturant concentration. However, little or no information is obtained about the actual structural nature of the unfolded conformations sampled in the denatured ensemble, or about how the distribution of populations of the various conformations in the denatured state depends on denaturant concentration.

In this paper we report a small-angle X-ray scattering (SAXS)[1] study of the equilibrium distribution of unfolded conformations of cytochrome *c* (Cyt *c*). Cyt *c* is a small protein (104 residues) with a heme group covalently bound to residues Cys14 and Cys17. The heme iron has two axial ligands: His18 and Met80 (*7*). Under typical denaturing conditions, such as high guanidine hydrochloride (GdnHCl) or urea concentrations at neutral pH, Met80 dissociates from the heme (*8*). The lack of conformational freedom of His18, which has two neighboring residues (Cys14, Cys17) covalently bound to the heme, prevents it from detaching from the heme iron even under denaturing conditions (*8*−*10*).

Equilibrium unfolding of Cyt *c* has been studied by many optical methods, including circular dichroism (CD) and tryptophan (Trp) fluorescence. CD at 222 nm, which

† This work was supported by National Institutes of Health Grant RR-01209 (to K.O.H.) and by National Science Foundation MCB95-07280 (to A.L.F.). Data were collected at Beam Line 4-2 at Stanford Synchrotron Radiation Laboratory (SSRL). SSRL is supported by the U.S. Department of Energy, the Office of Basic Energy Sciences, and in part by the National Institutes of Health, the National Center for Research Resources, the Biomedical Technology Program, and the Department of Energy, Office of Biological and Environmental Research.

* To whom correspondence should be addressed. Department of Applied Physics, Stanford University, Stanford, CA 94305. Tel: 650 723 4786. Fax: 650 723 2189. E-mail: doniach@drizzle.stanford.edu.

‡ Department of Physics, Stanford University.

§ University of California.

|| Department of Chemistry, Stanford University and Stanford Synchrotron Radiation Laboratory.

[1] Abbreviations: SAXS, small-angle X-ray scattering; GdnHCl, guanidine hydrochloride; SVD, singular value decomposition; R_g, radius of gyration; N, native; U_1, unfolded state 1; U_2, unfolded state 2; N/A, not applicable.

S0006-2960(98)00535-2 CCC: $15.00
Published on Web 08/20/1998

indicates the amount of secondary structure, shows a cooperative transition with an absence of such structure above 3.5 M GdnHCl (*11, 12*). A thermodynamic fit was performed on the data to yield C_m, the midpoint of the transition; ΔG°_U, the change in free energy of state U relative to the native state in the absence of denaturant; and *m*, the linear dependence of the free energy upon the denaturant concentration. This fit gives C_m, ΔG°_U, and *m* values of 2.7 M GdnHCl, 8.2 ± 0.8 kcal/mol, and 3.0 ± 0.3 kcal mol^{-1} M^{-1}, respectively (*12*). Trp fluorescence has long been used as a conformational probe for Cyt *c* (*13*−*18*). Energy transfer with the heme quenches Trp fluorescence. The measured degree of quenching can be used to calculate the average heme−Trp distance (*18*). Equilibrium unfolding Trp fluorescence studies show a cooperative transition with the transition complete upon reaching GdnHCl concentrations of 3.5 M (*13*−*17*). At higher GdnHCl concentrations, the fluorescence increases linearly; Chan et al. (*14*) attribute this to a continued expansion of the unfolded state. A thermodynamic fit to the Trp fluorescence data gives C_m, ΔG°_U, and *m* values of 2.77 ± 0.02 M GdnHCl, 9.7 ± 1.1 kcal/mol, and 3.5 ± 0.4 kcal mol^{-1} M^{-1} (*14*). While the ΔG°_U and *m* values measured by Trp fluorescence are higher than those measured by CD, the values from the two methods do agree within experimental error. Therefore, comparison of CD and Trp fluorescence data can only distinguish two states: one native and the other unfolded.

In the present SAXS study, a series of scattering intensity profiles, $I(S,c_k)$, where $S = (2 \sin \theta)/\lambda$ is the momentum transfer, 2θ is the scattering angle, and λ is the wavelength of the X-ray beam, were measured for a series of GdnHCl concentrations (denoted by c_k). By means of singular value decomposition (SVD), we are able to discern the minimum number of linearly independent functions needed to represent the entire series of scattering profiles (*19*). This allows us to gain information on the number of distinct unfolded components, and therefore the number of distinct equilibrium states of the protein needed to represent the denaturation equilibrium as a function of GdnHCl concentration. Within the limitations of the data quality accessible with the present experimental setup, we find that a minimum of three linearly independent functions is required to represent the data acquired over the range of denaturants measured, and hence that the ensemble of denatured configurations is made up of at least two (and quite possibly more) thermodynamically distinct states. Interestingly enough, we see very little change in the radius of gyration (determined with the very low angle data) of the unfolded state as the concentration of GdnHCl is increased, although inspection of the scattering profiles at high angles shows the degree of packing to decrease with increasing denaturant concentration. Thus our measurements show that the conformations populating the denatured subensemble at intermediate [GdnHCl] are more compact than would be expected under random coil conditions, which are assumed to pertain in the limit of high denaturant concentration.

MATERIALS AND METHODS

Materials. Horse heart cytochrome *c* was purchased from Sigma Chemical Company, St. Louis, MO. The protein was not purified further. Phosphate stock buffer solutions (100 mM) were prepared with (4.8 M) and without GdnHCl at pH 7. Stock buffers were filtered with a 0.2 μm acetate filter. Appropriate mixing ratios were used to achieve solutions of intermediate GdnHCl concentration: 0.6, 1.2, 1.8, 2.2, 2.4, 2.6, 2.8, 3.0, 3.2, 3.4, 3.6, 4.2, and 4.8 M. Protein was dissolved directly into these solutions; the sample was subsequently filtered with a 0.2 μm filter to remove undissolved protein from the sample. Cytochrome *c* concentrations were measured by UV absorption (ϵ_{280} = 8.55 mL mg^{-1} cm^{-1}.

SAXS. Measurements were made using the SAXS instrument on Beam Line 4-2 at Stanford Synchrotron Radiation Laboratory (*20*). X-ray energy was selected at 8980 eV (Cu edge) by a pair of Mo/B_4C multilayer monochromator crystals (*21*). Scattering patterns were recorded by a linear position-sensitive proportional counter, which was filled with an 80% Xe/20% CO_2 gas mixture. Scattering patterns were normalized by incident X-ray flux, which was measured with a short-length ion chamber before the sample. The sample-to-detector distance was calibrated to be 230 cm, using a cholesterol myristate sample. To avoid radiation damage of the sample, the protein solution was continuously passed through a 1.3 mm path length observation flow cell with 25 μm mica windows. The flow rate was ∼1.8 mL/s. Protein solution (1.5 mL) was used for each measurement and recirculated during the course of the measurement.

Guinier Analysis. Radii of gyration were calculated according to the Guinier approximation (*22*):

$$\ln(I(S)) = \ln(I(0)) - \frac{4\pi^2 R_g^{\,2}}{3} S^2 \qquad (1)$$

where R_g = radius of gyration, $S = (2 \sin \theta)/\lambda$, 2θ = scattering angle, and λ = X-ray wavelength. Guinier fits and errors were calculating using KaleidaGraph (Synergy Software, Reading, PA). The fitting range used was 0.0045−0.0071 Å^{-1} in *S*. Scattering profiles were taken on samples of protein concentrations of 5−6 mg/mL for Guinier analysis.

Kratky Data. To obtain good quality Kratky plots at high angles, high protein concentrations must be used. At each GdnHCl concentration, a scattering profile was first recorded at a high protein concentration of 20−24 mg/mL (profile A). However, at high protein concentrations, interparticle interference degrades the scattering profile at small angles (*23*). To reduce interparticle interference effects at small angles, the high protein concentration solution was diluted 4-fold to 5−6 mg/mL. A scattering profile was then recorded on this low protein concentration solution (profile B). The Kratky data at each [GdnHCl] were compiled in the following manner: (i) when $S < 0.013$, profile B was used; (ii) when $S > 0.018$, profile A multiplied by 0.25 was used; and (iii) in the region $0.013 < S < 0.018$, the profiles were merged as follows in order to smooth the joining region:

$$I(S) = 0.25 f_A I_A(S) + f_B I_B(S) \qquad (2)$$

where $f_A = (S - 0.013)/0.005$; $f_B = 1 - (S - 0.013)/0.005$; and I_A and I_B are profiles A and B, respectively. The 0.25 factor compensates for the fact that the protein sample for profile A was four times that for profile B. Profile B was used for Guinier analysis. Each merged scattering profile was normalized by the $I(0)$ value gained from the Guinier analysis.

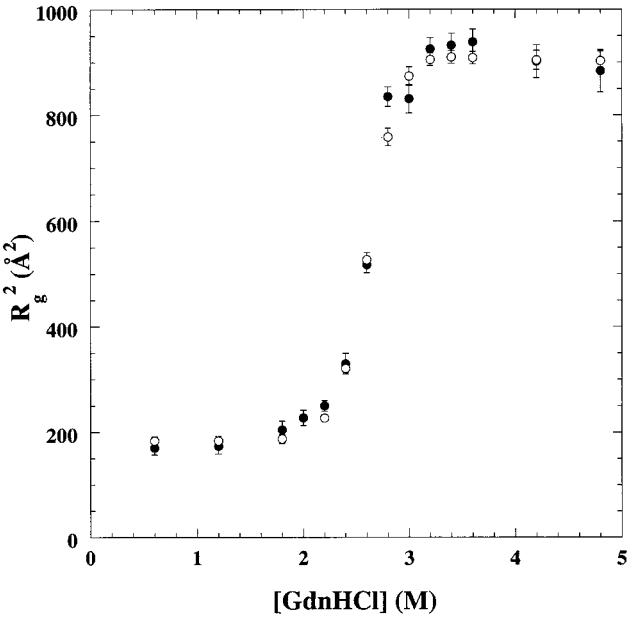

FIGURE 1: The radius of gyration (R_g) dependence upon GdnHCl at pH 7 (100 mM phosphate). R_g^2 displays a linear dependence upon the populations of different protein states (see eq 3). Therefore, R_g^2 vs [GdnHCl] is plotted here. The closed circles (●) were obtained from Guinier fits to the original SAXS data while the open circles (○) were obtained from Guinier fits to the scattering profiles reconstructed using singular value decomposition and a thermodynamic model. The same fitting range, 0.0045−0.0070 Å^{-1} in *S*, was used in all Guinier fits.

RESULTS

Guinier Data. The size of a protein can be determined from the X-ray scattering in the small-angle region. The protein scattering in this region, called the Guinier Region, is given as $I(S) = I(0)e^{-4\pi^2 R_g^2 S^2/3}$, where R_g is the protein's radius of gyration. Radii of gyration are obtained from the slope of the curve $\ln(I(S))$ vs S^2 in the Guinier Region (see Materials and Methods Section).

The unfolding of Cyt *c* monitored by the change in R_g shows a cooperative unfolding transition on addition of GdnHCl at pH 7 (Figure 1). In its native form, Cyt *c* has a R_g of 13.8 ± 0.3 Å. Above 3 M GdnHCl, R_g does not change, within error, and has a value of 30.3 ± 0.1 Å. The native value agrees with a previously published study (13.5 ± 0.1 Å), while the value of the denatured R_g is slightly lower than the value of 32.4 ± 1.6 Å found in the same study (*24*).

A two-state thermodynamic fit was performed on R_g^2, assuming a linear dependence of the free energy, ΔG ($\Delta G_U = \Delta G_U^\circ - m[\text{GdnHCl}] = m(C_m - [\text{GdnHCl}])$ (*25*). For a two-state system, the measured R_g is an average radius of gyration of all species in solution:

$$R_g^2 = f_N R_N^2 + f_U R_U^2 \quad (3)$$

where f_N = native fraction, f_U = unfolded fraction, R_N = radius of gyration for the native state, R_U = radius of gyration for the unfolded state, and neglecting contrast changes which are small in the present case (*19*). Since R_g^2 maintains a linear dependence upon the fractional populations, R_g^2 is the relevant quantity to fit (not R_g). The C_m and *m* values were found to be 2.6 ± 0.1 M GdnHCl and 4.5 ± 0.7 kcal mol^{-1} M^{-1}, respectively. While the C_m value of CD and R_g^2 data agree within error, the *m* value obtained from the R_g^2 curve is significantly larger than the CD value (3.0 ± 0.3 kcal mol^{-1} M^{-1} (*12*). This suggests that the unfolding transition perhaps cannot simply be described as a two-state transition.

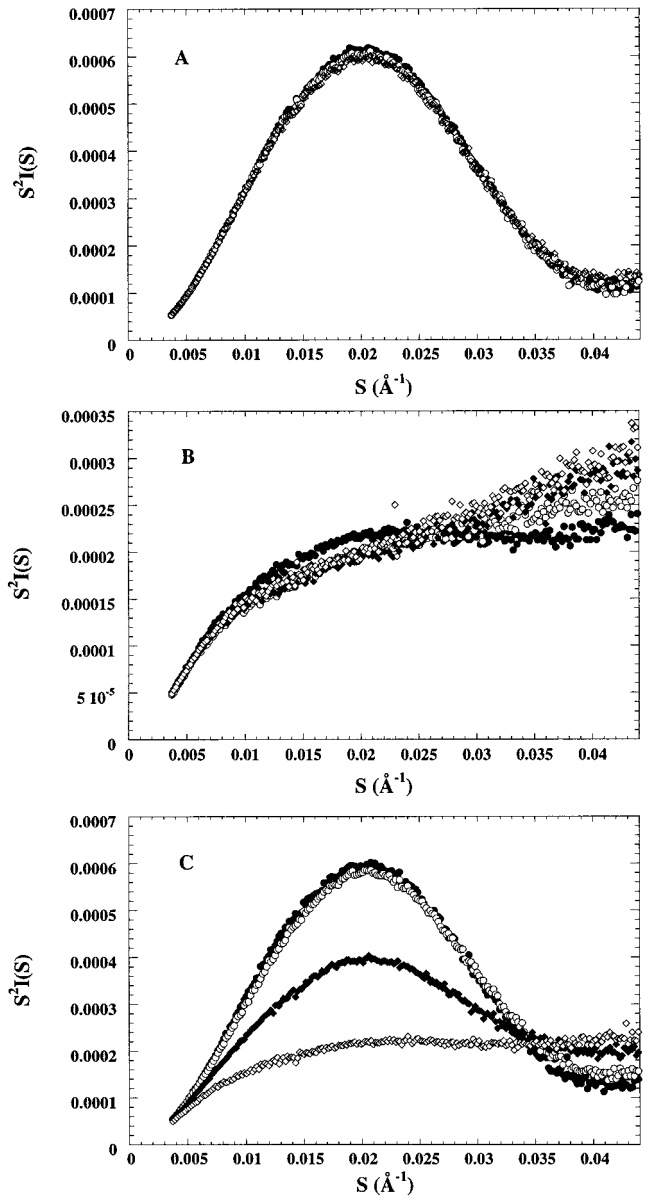

FIGURE 2: Kratky plots, $S^2I(S)$ vs *S*, of Cyt *c* in (A) native conditions: 0.6 (●), 1.2 (○), and 1.8 M (◇) GdnHCl; (B) denaturing conditions: 3.0 (●), 3.6 (○), 4.2 (◆), and 4.8 M (◇) GdnHCl; (C) cooperative unfolding region: 1.8 (●), 2.2 (○), 2.6 (◆), and 3.0 M (◇) GdnHCl.

Kratky Data. Additional structural information can be obtained from the SAXS data by inspecting the entire scattering pattern. The degree of unfolding is best gauged by viewing a Kratky plot, $S^2I(S)$ vs *S*, which emphasizes the *S* region between 0.02 and 0.04 Å^{-1}, which we will refer to as the high-angle part of the scattering profile. Globular proteins scatter as S^{-4} (*26*) at high *S* values, yielding a Kratky plot that is proportional to S^{-2}. Such dependence is shown in the Kratky plots of Cyt *c* in 0.6, 1.2, and 1.8 M GdnHCl at high angles (Figure 2A). The fact that identical Kratky plots were determined under these three conditions, complemented with the nearly identical radii of gyration, demonstrates that, as expected, Cyt *c* is fully native at these denaturant concentrations.

A random coil scatters as S^{-1} at high angles (*22*). A Kratky plot of a random coil then shows an S^1 dependence at these angles. From the Guinier analysis, Cyt *c* appears to be fully unfolded above 3 M GdnHCl as R_g does not change (~30 Å) (Figure 1). However, Kratky plots reveal changes in the protein structure at these denaturant concentrations (Figure 2B). For 3 M GdnHCl, the Kratky plot is flat at high *S* (i.e., at high angles) as seen from the curve of filled circles (●) on this figure. With increasing denaturant concentration, an increasing tendency toward linearity is seen at these angles. These changes indicate an increasing random coil content of the protein sample. Thus, despite the large radius of gyration, the Kratky plots imply that residual structure exists at 3 M GdnHCl. These results indicate that in 3 M GdnHCl the protein assumes conformations more tightly packed than a random coil, and that the observed R_g is rather insensitive to these changes in packing density.

Corroboration of residual structure still existing at GdnHCl concentrations where R_g has saturated comes from circular dichroism experiments. For example, under the conditions used in the SAXS experiments, at 2.8 M GdnHCl CD_{222} measurements of Cyt *c* indicate significantly more structure than the unfolded protein, namely ~40% of the signal of the native structure (*12*). Even at 3.0 M GdnHCl there is still significant structure remaining by CD, and in fact, the CD transition is not complete until close to 3.5 M GdnHCl (*12*).

The greatest conformational changes can be observed by studying the Kratky plots in the transition region, 2.0−3.0 M GdnHCl (Figure 2C). Here the peak magnitude of the plot is seen to decrease as the [GdnHCl] is increased, suggesting loss of globularity in the protein sample (*24*). Further, the Kratky plots move up in the high-angle region, suggesting loss of structure and thus an increase in random coil content.

For a purely two-state system, an isoscattering point would be expected in the Kratky plots (*24*). A scattering profile at any denaturant concentration (c_k) can be expressed as a linear superposition of the native and denatured scattering profiles:

$$I(S,c_k) = f^k_N I_N(S) + f^k_U I_U(S) \quad (4)$$

The scattering patterns of the native and denatured states intersect ($I_N(S_0) = I_U(S_0)$, where S_0 is the isoscattering point). Since $f_N + f_U = 1$, the scattering profile at any denaturant concentration will always have the same value at S_0. However, Figure 2 shows that an isoscattering point does not exist for the entire [GdnHCl] series. This suggests that more than two equilibrium states of Cyt *c* are populated at pH 7.

Singular Value Decomposition. Since inspection of the Kratky plots suggests that more than two equilibrium states exist, the singular value decomposition (SVD) method was employed to identify the number of states. Unlike the Guinier (R_g) analysis, which uses only the small-angle region, the SVD method was applied to the entire scattering profile. SVD deconvolutes the scattering patterns of the [GdnHCl] series into an orthogonal series of basis scattering curves. A scattering pattern at any [GdnHCl] can be constructed as a linear superposition of these basis scattering curves.

The scattering profiles at all [GdnHCl] can be compiled into a matrix, $A(S,k)$, where a column of matrix *A* represents the scattering profile at a specific [GdnHCl] denoted by the index *k*: $A(S,k) = I(S,c_k)$. If there are *M* points (*S* values) in each scattering profile and *P* GdnHCl concentrations, then *A* is an $M \times P$ matrix. SVD can be used to decompose matrix *A* into the product of three matrixes:

$$A(S,k) = u(S,k)s\upsilon^T \quad (5)$$

The columns of $u(S,k)$, an $M \times P$ matrix, are orthogonal basis scattering curves (Figure 3). Scattering profiles are a linear superposition of these basis scattering curves. The diagonal components of *s*, a $P \times P$ matrix, are called the singular values (w_i). The product $s\upsilon^T$ determines the [GdnHCl]-dependent weight applied to the individual basis scattering curves when constructing the scattering profiles. The scattering profile at GdnHCl concentration, c_k, can then be approximated:

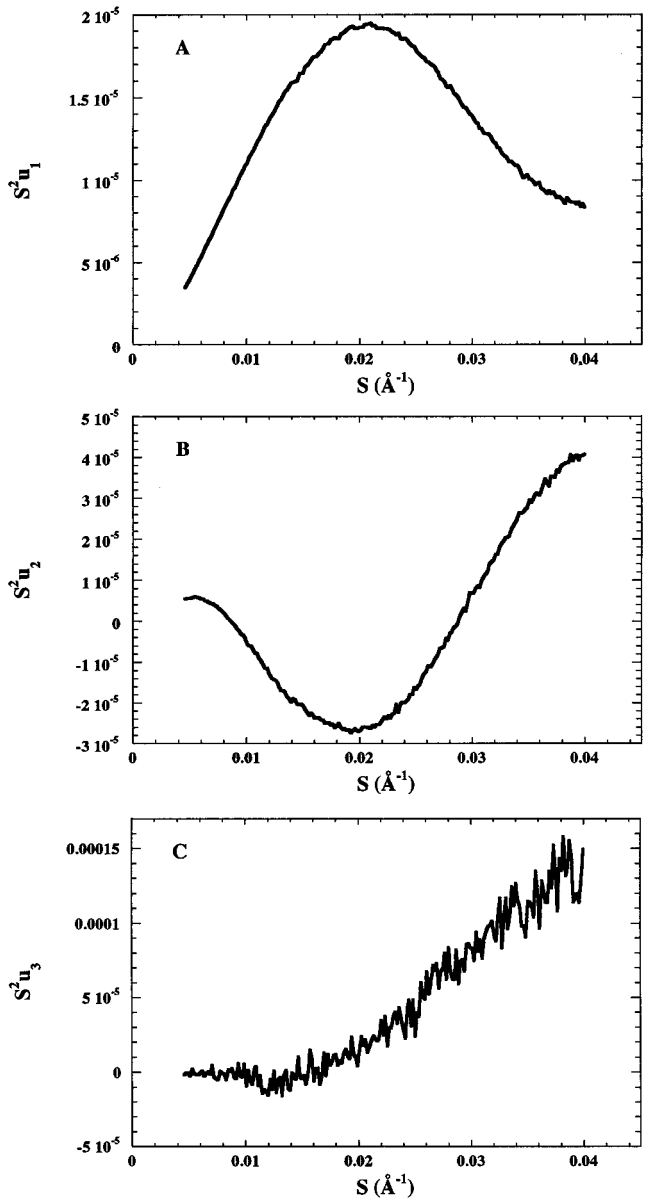

FIGURE 3: Kratky plots of the singular value decomposition (SVD) basis scattering curves (columns of matrix *u*): (A) S^2u_1; (B) S^2u_2; and (C) S^2u_3.

$$I(S,c_k) = \sum_{j=1}^{L} w_j v_{kj} u_j(S) = \sum_{j=1}^{L} w_j b_j^k u_j(S) \quad (6)$$

where L is the minimum number of components necessary to adequately reconstruct the scattering profiles and $b_j^k = v_{kj}$ is the basis coefficient j at GdnHCl concentration, c_k. Terms with $j > L$ represent noise. Because of the linear relationship of the SVD fits, the variance of the fitting coefficients, $w_j b_j^k$, is a linear relationship of the individual data points $I(S,k)$:

$$\sigma_j^2 = \sum_s \sigma_s^2(k) u_j^2(S) \quad (7)$$

where $\sigma_S^2(k)$ is the variance in $I(S,k)$ and the summation is over all S values (*27*). Chen et al. (*19*) have applied the SVD method to SAXS spectra.

Each SAXS profile is a linear superposition of the profiles of the distinct states that can be populated by the protein (for a two-state system, see eq 4). For the states to be distinct, their scattering profiles must be linearly independent. The number of states therefore indicates the number of linearly independent scattering curves needed to construct the scattering profile at any [GdnHCl]. The number of significant SVD components (L) determines the minimum number of linearly independent scattering profiles needed to construct the series of scattering patterns, $I(S,c_k)$. Therefore, the number of significant SVD components reveals the number of distinguishable subensembles of conformations or, in other words, the number of thermodynamic states. Four factors were considered in determining the number of significant SVD components: (1) singular values; (2) autocorrelations of u; (3) shape of basis functions; and (4) reduced chi-squared of SVD reconstruction of the scattering profiles (*19*, *28*). An analysis using these factors shows at least three components to be significant and a fourth component possibly being significant as well (Figures 3 and 4; data of fourth component not shown).

Due to the mathematical properties of SVD, the basis scattering curves do not represent scattering patterns of real conformations (*28*). This is evidenced by u_2 dipping below 0. (Kratky plots of the basis scattering curves are presented in Figure 3.) However, they can illustrate the structural changes seen in the protein. The Kratky plot of the first basis function, u_1, is characteristic of a globular protein (Figure 3A). This basis function is most heavily weighted at low [GdnHCl], where the protein is native (Figure 4A). The second component is negative at intermediate angles and positive and increasing at high angles (Figure 3B). Positive weighting of this component would have the following effects on a Kratky plot: (i) decrease the peak magnitude, signifying reduced globularity of the protein and (ii) cause an increase at high angle, indicative of increasing random coil content. The third component increases linearly at high angles (Figure 3C). Positive weighting of this component would lead to an upward sloping Kratky plot at high angle, thus signifying an increase in the random coil content of the protein sample. This component's negligible contribution at small and intermediate angles means that it would have no effect on the peak magnitude of the overall Kratky plot.

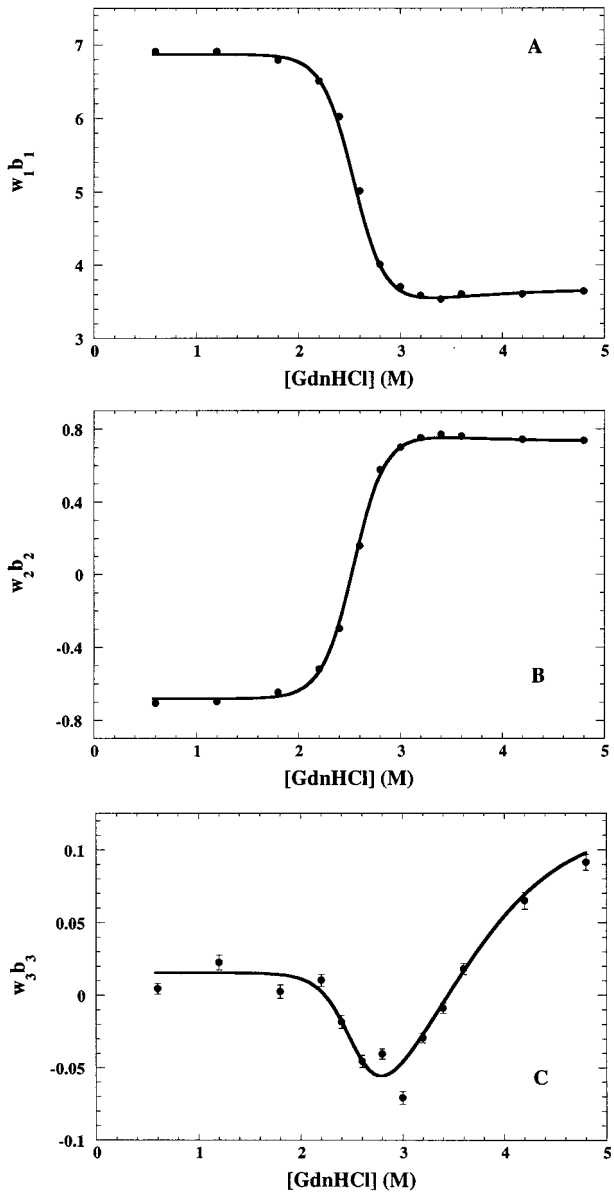

FIGURE 4: Basis function coefficients (columns of matrix $b = v^T$) as a function of denaturant concentration. The curves are coefficients for basis functions: (A) u_1; (B) u_2; (C) and u_3. The solid lines represent the fits resulting from the thermodynamic model.

Thermodynamic Model. Just as the scattering profiles at any [GdnHCl], $I(S,c_k)$, can be represented by a linear superposition of the basis scattering curves, so can the scattering profiles of the individual thermodynamic states (labeled N, U_1, and U_2):

$$I_N(S) = w_1 b_1^N u_1(S) + w_2 b_2^N u_2(S) + w_3 b_3^N u_3(S) \quad (8a)$$

$$I_{U1}(S) = w_1 b_1^{U1} u_1(S) + w_2 b_2^{U1} u_2(S) + w_3 b_3^{U1} u_3(S) \quad (8b)$$

$$I_{U2}(S) = w_1 b_1^{U2} u_1(S) + w_2 b_2^{U2} u_2(S) + w_3 b_3^{U2} u_3(S) \quad (8c)$$

where $I_N(S)$, $I_{U1}(S)$, and $I_{U2}(S)$ are the scattering profiles for N, U_1, and U_2, respectively; w_i are the singular values; b_i^N, b_i^{U1}, and b_i^{U2} are the basis coefficients for N, U_1, and U_2, respectively; and $u_1(S)$, $u_2(S)$, and $u_3(S)$ are the basis scattering curves. Using a minimal thermodynamic model, we determined the proper linear combinations of the basis curves required to represent the scattering profiles of the

distinct thermodynamic states. In other words, the thermodynamic model determined b_i^N, b_i^{U1}, and b_i^{U2}, where $i = 1, 2, 3$. Each scattering profile in the [GdnHCl] series can be thought of as the superposition of the scattering profiles for these three thermodynamic states:

$$I(S,c_k) = f^k_N I_N(S) + f^k_{U1} I_{U1}(S) + f^k_{U2} I_{U2}(S) \quad (9)$$

where f^k_N, f^k_{U1}, and f^k_{U2} are the fractional populations of states N, U_1, and U_2, respectively, at GdnHCl concentration c_k; $f^k_N + f^k_{U1} + f^k_{U2} = 1$; and $I_N(S)$, $I_{U1}(S)$, and $I_{U2}(S)$ are the scattering patterns for N, U_1, and U_2, respectively. The fractional populations of each state can be expressed

$$f^k_N = \frac{1}{1 + e^{-\Delta G_{U1}/RT} + e^{-\Delta G_{U2}/RT}} \quad (10a)$$

$$f^k_{U1} = \frac{e^{-\Delta G_{U1}/RT}}{1 + e^{-\Delta G_{U1}/RT} + e^{-\Delta G_{U2}/RT}} \quad (10b)$$

$$f^k_{U2} = \frac{e^{-\Delta G_{U2}/RT}}{1 + e^{-\Delta G_{U1}/RT} + e^{-\Delta G_{U2}/RT}} \quad (10c)$$

where ΔG_U is the change in free energy of state U relative to the native state. In our model, the free energy was assumed to have a linear dependence on the denaturant concentration (*25*)

$$\Delta G_{U1} = \Delta G^0_{U1} - m_1[\text{GdnHCl}] \quad (11a)$$

$$\Delta G_{U2} = \Delta G^0_{U2} - m_2[\text{GdnHCl}] \quad (11b)$$

where ΔG^0_U is the change in free energy of state U relative to the native state at neutral pH and in the absence of denaturants, and m is the linear dependence of ΔG_U upon the denaturant concentration.

Using eq 8 to substitute for $I_N(S)$, $I_{U1}(S)$, and $I_{U2}(S)$ in eq 9, we find

$$\begin{aligned} I(S,c_k) = {} & f^k_N(w_1 b_1^N u_1(S) + w_2 b_2^N u_2(S) + w_3 b_3^N u_3(S)) + \\ & f^k_{U1}(w_1 b_1^{U1} u_1(S) + w_2 b_2^{U1} u_2(S) + w_3 b_3^{U1} u_3(S)) + \\ & f^k_{U2}(w_1 b_1^{U2} u_1(S) + w_2 b_2^{U2} u_2(S) + w_3 b_3^{U2} u_3(S)) = \\ & (f^k_N b_1^N + f^k_{U1} b_1^{U1} + f^k_{U2} b_1^{U2}) w_1 u_1(S) + (f^k_N b_2^N + \\ & f^k_{U1} b_2^{U1} + f^k_{U2} b_2^{U2}) w_2 u_2(S) + (f^k_N b_3^N + f^k_{U1} b_3^{U1} + \\ & f^k_{U2} b_3^{U2}) w_3 u_3(S) \quad (12) \end{aligned}$$

For a three-state model, comparison of eq 6 (with $L = 3$) and eq 12 gives the following relations:

$$b^k_1 = f^k_N b_1^N + f^k_{U1} b_1^{U1} + f^k_{U2} b_1^{U2} \quad (13a)$$

$$b^k_2 = f^k_N b_2^N + f^k_{U1} b_2^{U1} + f^k_{U2} b_2^{U2} \quad (13b)$$

$$b^k_3 = f^k_N b_3^N + f^k_{U1} b_3^{U1} + f^k_{U2} b_3^{U2} \quad (13c)$$

The thermodynamic parameters (ΔG^0_{U1}, m_1, ΔG^0_{U2}, m_2) and the basis coefficients for the thermodynamic states (b_i^N, b_i^{U1}, b_i^{U2} where $i = 1, 2, 3$) can then be identified by minimizing:

$$\chi^2 = \sum_{j=1}^{L} \sum_{k=1}^{P} \frac{w_j^2 (b^k_{j,\text{obs}} - b^k_{j,\text{cal}})^2}{\sigma_j^2(k)} \quad (14)$$

where L is the number of relevant SVD components; P is the number of scattering profiles recorded (i.e., the number of [GdnHCl]); $b^k_{j,\text{obs}}$ are the basis coefficients output from the SVD analysis; $b^k_{j,\text{cal}}$ are the values calculated for the basis functions using eq 13; and $\sigma_j^2(k)$ are the variances in the SVD weighting coefficients. (Equivalently, the thermodynamic model could be used to refine the data directly against the SVD-defined U_1 and U_2 profiles.)

Both two-state (N ↔ U) and three-state (N, U_1, U_2) thermodynamic models were employed to fit the SVD weighting coefficients. The thermodynamic parameters resulting from the fits are presented in Table 1. The three-state fits to the basis function coefficients, b_i, are plotted as solid lines in Figure 4.

The populations of the three states as a function of denaturant concentration (calculated using eq 10) are presented in Figure 5. U_1 reaches a maximum population of ~60% at 2.9 M GdnHCl. Due to its large energy barrier (ΔG^0_{U2}), U_2 does not become well populated until high [GdnHCl] is reached.

As mentioned earlier, Trp fluorescence has long been used as a conformation probe for Cyt *c* (*13*−*18*). Since Trp fluorescence indicates conformational changes, it would provide a good comparison to a SAXS study. The equilibrium unfolding of Cyt *c* at pH 7 has been followed by Trp fluorescence (*15*). The thermodynamic parameters from a two-state (N ↔ U) fit were found to be $\Delta G^0_U = 9.7 \pm 1.1$ kcal/mol and $m = 3.5 \pm 0.4$ kcal mol^{-1} M^{-1}. ΔG^0_U from our two-state fit (10.8 kcal/mol) is slightly higher than that found in the Trp fluorescence study. ΔG^0_{U1} from the three-state fit (9.1 kcal/mol) is in better agreement and m_1 (3.4 kcal mol^{-1} M^{-1} is nearly identical to the fluorescence m value, suggesting the cooperative transition seen by Trp fluorescence is the N → U_1 transition.

Constructed States. With the results of the thermodynamic fit to the basis function coefficients, scattering profiles for the three distinct thermodynamic states (N, U_1, and U_2) can be constructed using eq 8. Guinier analysis of these profiles yields radii of gyration of 13.4 ± 0.2, 29.9 ± 0.5, and 29.6 ± 0.3 Å for N, U_1, and U_2, respectively. Kratky plots of the constructed scattering profiles are presented in Figure 6. From the plot, N is clearly globular and nearly identical to the original Kratky plots at 0.6, 1.2, and 1.8 M GdnHCl. The differences between U_1 and U_2 are evident at high angles. U_1 is flat at large angles while U_2 continues to increase linearly. The Kratky plot of U_2 strongly resembles that of a random coil, indicating a complete lack of structure. The plateau in the Kratky plot of U_1 suggests evidence of some residual structure within the protein. The absence of a peak at small angles demonstrates the lack of globularity in U_1 and U_2.

The m_i value is roughly proportional to the difference in solvent-accessible surface area between the native state (N) and the unfolded state (U_i) (*29*, *30*). As structure develops within the protein, some residues will become protected from the solvent, presumably resulting in a reduction of the solvent-accessible surface area as compared to the completely

Table 1: Thermodynamic Parameters Resulting from Fits to the SVD Components

model	ΔG^0_{U1} (kcal/mol)	m_1 (kcal mol^{-1} M^{-1})	ΔG^0_{U2} (kcal/mol)	m_2 (kcal mol^{-1} M^{-1})
two-state	10.8	4.2	N/A	N/A
three-state	9.4	3.6	12.7	4.6

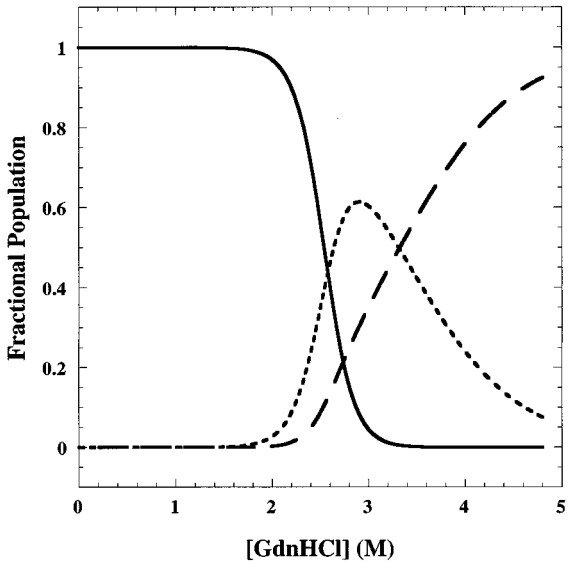

FIGURE 5: Populations of three states from the thermodynamic model as a function of denaturant concentration: N (solid line), U_1 (dashed line), and U_2 (broken line).

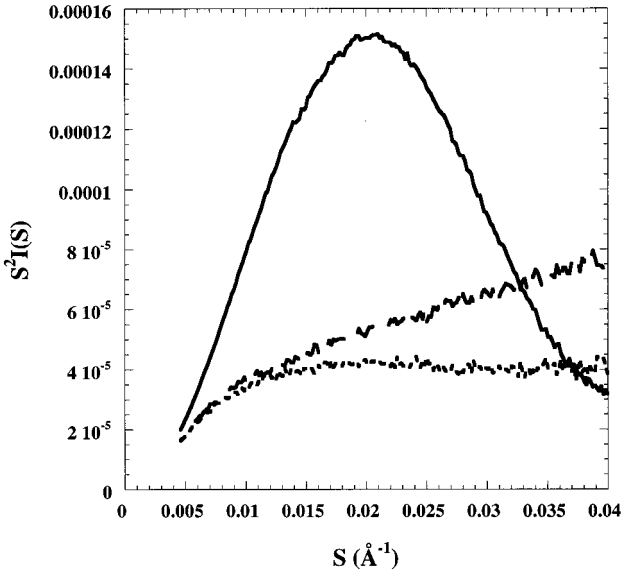

FIGURE 6: Scattering profiles of three states generated from the thermodynamic model: N (solid line), U_1 (dashed line), and U_2 (broken line).

unfolded state. With this view, the lower *m* value of U_1 as compared to U_2 would be consistent with the presence of residual structure in U_1.

Lack of Interparticle Scattering Effects. At the protein concentrations used for SAXS measurements, protein aggregation can be a concern. To ensure that U_1 is not an aggregated state, we took SAXS spectra at different protein concentrations in 3.0 M GdnHCl, a [GdnHCl] close to the peak U_1 population. If association were occurring, one would expect an increase in protein concentration to lead to an increase in R_g as more of the sample is aggregating. However, a decrease in R_g was observed with increasing concentration (~1 Å from 4 to 20 mg/mL), which is expected for a monomeric protein. The decrease in R_g is a result of interparticle interference (*23*).

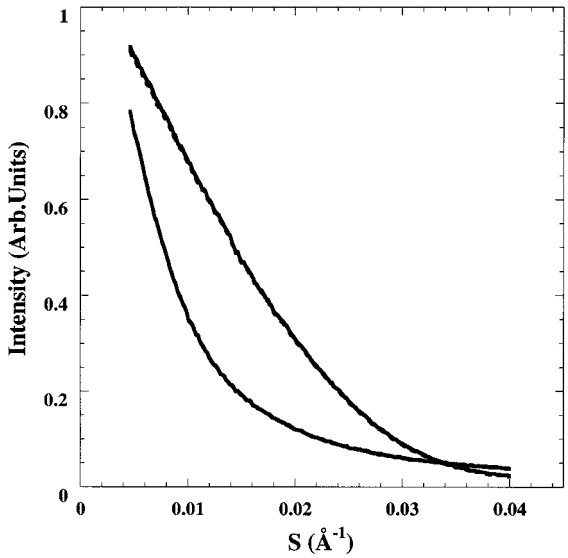

FIGURE 7: Reconstructed scattering profiles at 2.4 and 3.6 M GdnHCl. Solid lines are scattering profiles from the 3 component SVD reconstruction. Dashed lines are scattering profiles recreated using the parameters of the thermodynamic model. The upper curve is 2.4 M GdnHCl, and the lower curve is 3.6 M GdnHCl.

Additional evidence for lack of aggregation is seen in the SVD components. SVD analysis of two proteins (lysozyme and Cyt *c*) has shown a third basis function which only varies in the small-angle region ($S \leq 0.01$ Å^{-1}) where interparticle scattering is important (unpublished data). This is not the case for the present data. The third component does not contribute to the scattering profile in the small-angle region (Figure 3C).

Reconstructed Profiles. The Cyt *c* scattering profiles at varying [GdnHCl] can be reconstructed using the scattering profiles of our three model states: N, U_1, and U_2 (eq 9). A good test of our three-state model is to compare these reconstructed profiles with the three SVD component reconstructed profiles. The reconstructions of the 2.4 and 3.6 M GdnHCl profiles are presented in Figure 7.

The reconstructed scattering profiles were fitted in the Guinier region for R_g. These values are in good agreement with the radii of gyration calculated from the original SAXS profiles (Figure 1). As mentioned earlier, initial inspection of the R_g^2 curve suggests a two-state transition with increasing [GdnHCl]. However, our three-state model shows U_1 and U_2 to have nearly identical radii of gyration (30 Å), and thus, the two states would not be distinguishable in a R_g^2 vs [GdnHCl] curve.

DISCUSSION

Our SAXS studies provide evidence for more than one denatured subensemble of Cyt *c* at neutral pH. SVD analysis combined with a thermodynamic model has identified at least two denatured states. These states have similar radii of gyration but differ in their degree of compactness. U_2 has a Kratky plot characteristic of a random coil. Meanwhile,

the plateau of the Kratky plot of U_1 at high angles indicates the existence of some residual structure in U_1.

If U_1 has structure, what type of structure is it? The best way to address this question is to compare our results with studies which have identified partially folded states of Cyt *c* in equilibrium. Cyt *c* unfolds at pH 2, with the addition of salt leading to a partially folded state (labeled the "A-state"), which has a significant amount of secondary structure but a disordered tertiary structure (*12*, *31*−*34*). Two-dimensional NMR hydrogen exchange reveals that the three major helices of Cyt *c* remain intact in the A-state (*35*). This state has been studied by SAXS, with its R_g measured to be 17.4 Å (*12*, *24*). The A-state therefore is fairly compact as compared with an unfolded R_g of 30 Å. This is further demonstrated by the globular nature of the A-state's Kratky plot (*24*). Conversely, U_1 has an expanded R_g with a Kratky plot displaying no evidence of globularity. By comparison with the A-state, it is clear that U_1 cannot possess significant secondary structure.

Very small populations of partially folded states have been identified by hydrogen exchange under native conditions (*36*). Bai et al. exposed the sequential unfolding of structural subunits with this method. In the least stable and least structured partially folded state (I_{NC}), the amino- and carboxyl-terminal helices exist as the only secondary structural units. These two helices cooperatively unfold together in the transition to the completely unfolded state. In the native state, these two helices are docked against each other, with their axes perpendicular (*7*). This interaction is believed to be preserved in the A-state (*35*) as well as in I_{NC} (*36*). Despite its probable lack of tertiary structure, the overall expansion of I_{NC} is limited by having both ends of the backbone bound together. Therefore, one would expect a smaller R_g than the completely unfolded state. U_1, however, has a similar R_g as the completely unfolded state (U_2) and therefore I_{NC} cannot serve as a model for U_1. As I_{NC} is the partially folded state with the least amount of secondary structure, U_1 appears to lack any stable secondary structure.

Dill and Shortle (*37*) have proposed that "weakly denatured proteins, either by temperature or by low denaturant concentrations, have considerable number of contacts among hydrophobic residues." These contacts can be non-native (*38*). Therefore, we propose that the structure evident in the U_1 state is nonspecific hydrophobic interactions. While the specific process in which GdnHCl acts as a denaturant is not known, it is known that GdnHCl weakens hydrophobic interactions (*30*, *39*). Therefore, as the GdnHCl concentration increases, the number of hydrophobic interactions decreases, allowing the protein to achieve more of a random coil structure. This would explain the lack of evidence of any structure in U_2, which is maximally populated at high GdnHCl concentrations.

The relatively large free energy difference between the U_1 and U_2 states points to a further consideration: proteins are charged heteropolymers, therefore they may be expected locally to experience inter-residue attractive forces in some conformations, even in "good solvent", that is, moderate denaturing conditions. Therefore the free energy landscape may be expected to have a multiplicity of local minima accessible under mildly denaturing conditions. Our observation of the U_1 state may thus be an indication of one of these local minima being populated as the denaturant concentration is lowered.

The fourth SVD component suggests the possibility that more than two denatured states may exist (data not shown); the poor resolution of the basis function coefficients made it too difficult to include a fourth state into our thermodynamic model. Shortle (*38*) has proposed that the unfolded state is "an ensemble of many poorly structured conformations or microstates that behaves as a single, more-or-less continuous distribution of microstates." Breaking of the non-native hydrophobic contacts is presumably not cooperative, but instead gradual. This gradual breaking of these contacts would lead to a multiplicity of unfolded states. Taking into consideration the effects of Coulombic interactions may modulate this continuum to some extent.

A recent study of Cyt *c* Trp fluorescence suggested a swelling of the unfolded state with increasing denaturant concentration. At high GdnHCl concentrations (above the cooperative transition), the fluorescence increases linearly. Normally, this has been believed to be solely a solvent effect on the fluorescence of the solvent-exposed Trp, rather than an indication of conformational change. Chan et al. (*14*) examined the fluorescence of a Cyt *c* fragment spanning residues 55−63, which contains the fluorescent residue Trp59, as a function of [GdnHCl]. Presumably the protein fragment and the holoprotein should display the same linear dependence. The fluorescence of this protein fragment displays a linear dependence on [GdnHCl]. However, the linear dependencies of the protein fragment and the holoprotein are not the same: the holoprotein displays a stronger linear dependence than the protein fragment. Chan et al. believe this suggests "expansion of the unfolded state (increase in the average heme−tryptophan distance) as the quality of the solvent improves." A nonsigmoidal increase in a physical quantity, in this case the Trp fluorescence, is consistent with a multiplicity of accessible denatured states.

We believe that our SAXS results are consistent with the existence of a multiplicity of denatured states. As mentioned before, SVD analysis of the scattering profiles does suggest the existence of at least one other denatured state. Presumably, with an increasing number of scattering profiles used for SVD, an increasing number of unfolded states would be identified. Therefore, we do not believe U_1 and U_2 to be the only unfolded states. In fact, these states probably represent averages of many unfolded states. U_1 is an average of the denatured states populated at mild denaturing conditions. U_2 is an average of the denatured states populated at strong denaturing conditions.

Our findings for Cyt *c* are also consistent with the observation of a partially folded intermediate for lysozyme at pH 2.9 by Chen et al. (*19*). In that case the intermediate was partially collapsed relative to the denatured state. Ibarra et al. (*40*) have recently pointed out that the data of Chen et al. could also be fit by a two-state model, provided the "baseline" for the R_g^2 fits could be allowed to depend on denaturant concentration (Sanchez-Ruiz, personal communication). This way of representing the data is in fact equivalent to allowing for a multiplicity of denatured states, since the only way R_g could depend on denaturant concentration is for there to be a multiplicity of denatured states whose populations vary as the denaturant concentration is changed.

Kinetic studies of the initial stages of refolding of Cyt *c* show the existence of a "burst phase", or rapid collapse within the dead time of the mixer used to initiate the refolding process (*14*, *15*, *17*, *41*). Chan et al. (*14*) and Sosnick et al. (*41*) suggest that the contraction is a downhill (barrier-free) process which can occur within 1 ms. Since our data does not show an appreciable change of R_g for the U_1 state, there are clearly further denatured states which become populated at still lower denaturant concentrations in these kinetic experiments. It is possible that more accurate SAXS studies at lower [GdnHCl] could detect the population of such more compact denatured states, which are also inferred from the fluorescence data of Sosnick et al.

ACKNOWLEDGMENT

We thank Dr. Hirotsugu Tsuruta for help with the SAXS setup on Beam Line 4-2 at SSRL. We thank Dr. James Hofrichter for very helpful discussions regarding the SVD method.

REFERENCES

1. Anfinsen, C. B. (1973) *Science 181*, 223−230.
2. Bax, A. (1989) *Annu. Rev. Biochem. 58*, 223−256.
3. James, T. L., and Basus, V. J. (1991) *Annu. Rev. Phys. Chem. 42*, 501−542.
4. Brünger, A. T. (1997) *Nat. Struct. Biol. 4*, 862−865.
5. Gillespie, J. R., and Shortle, D. (1997) *J. Mol. Biol. 268*, 158−169.
6. Wang, Y., and Shortle, D. (1997) *Folding Des. 2*, 93−100.
7. Bushnell, G. W., Louis, G. V., and Brayer, G. D. (1990) *J. Mol. Biol. 214*, 585−596.
8. Babul, J., and Stellwagen, E. (1971) *Biopolymers 11*, 2359−2361.
9. Babul, J., and Stellwagen, E. (1972) *Biochemistry 11*, 1195−1200.
10. Tsong, T. Y. (1975) *Biochemistry 14*, 1542−1547.
11. Tsong, T. Y. (1974) *J. Biol. Chem. 249*, 1988−1990.
12. Hamada, D., Kuroda, Y., Kataoka, M., Aimoto, S., Yoshimura, T., and Goto, Y. (1996) *J. Mol. Biol. 256*, 172−186.
13. Brems, D. N., and Stellwagen, E. (1983) *J. Biol. Chem. 258*, 3655−3660.
14. Chan, C.-K., Hu, Y., Takahashi, S., Rousseau, D. L., Eaton, W. A., and Hofrichter, J. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 1779−1784.
15. Colon, W., Elove, G. A., Wakem, L. P., Sherman, F., and Roder, H. (1996) *Biochemistry 35*, 5538−5549.
16. Elove, G. A., Chaffote, A. F., Roder, H., and Goldberg, M. E. (1992) *Biochemistry 31*, 6876−6883.
17. Elove, G. A., Bhuyan, A. K., and Roder, H. (1994) *Biochemistry 33*, 6925−6935.
18. Tsong, T. Y. (1976) *Biochemistry 15*, 5467−5473.
19. Chen, L., Hodgson, K. O., and Doniach, S. (1996) *J. Mol. Biol. 261*, 658−671.
20. Wakatsuki, S., Hodgson, K. O., Eliezer, D., Rice, M., Hubbard, S., Gillis, N., and Doniach, S. (1992) *Rev. Sci. Instrum. 63*, 1736−1740.
21. Tsuruta, H., Brennan, S., Rek, Z. U., Irving, T. C., Tompkins, W. H., and Hodgson, K. O. (1998) *J. Appl. Crystallogr.* (in Press).
22. Glatter, O., and Kratky, O. (1982) *Small Angle X-ray Scattering*, Academic Press, New York.
23. Hubbard, S. R., and Doniach, S. (1988) *J. Appl. Crystallogr. 21*, 953−959.
24. Kataoka, M., Hagihara, Y., Mihara, K. I., and Goto, Y. (1993) *J. Mol. Biol.* 591−596.
25. Pace, C. N. (1986) in *Methods in Enzymology* (Colowick, S. P., and Kaplan, N. O., Eds.) pp 266−280, Academic Press, Orlando, FL.
26. Porod, G. (1951) *Kolloid-Z. 124*, 83.
27. Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T. (1989) *Numerical Recipes: The Art of Scientific Computing*, Cambridge University Press, Cambridge, U.K.
28. Henry, E. R., and Hofrichter, J. (1992) *Methods Enzymol. 210*, 129−192.
29. Alonso, D. O. V., and Dill, K. A. (1991) *Biochemistry 30*, 5974−5985.
30. Schellman, J. A. (1978) *Biopolymers 17*, 1305−1322.
31. Goto, Y., Calciano, L. C., and Fink, A. L. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 573−577.
32. Goto, Y., Takahashi, N., and Fink, A. L. (1990) *Biochemistry 29*, 3480−3488.
33. Hagihara, Y., Tan, Y., and Goto, Y. (1994) *J. Mol. Biol. 237*, 336−348.
34. Ohgushi, M., and Wada, A. (1983) *FEBS Lett. 164*, 21−24.
35. Jeng, M.-F., Englander, S. W., and Elove, G. A. (1990) *Biochemistry 29*, 10433−10437.
36. Bai, Y., Sosnick, T. R., Mayne, L., and Englander, S. W. (1995) *Science 269*, 192−197.
37. Dill, K. A., and Shortle, D. (1991) *Annu. Rev. Biochem. 60*, 795−825.
38. Shortle, D. (1996) *FASEB J. 10*, 27−34.
39. Lee, J. C., and Timasheff, S. N. (1974) *Biochemistry 13*, 257−265.
40. Ibarra-Molero, B., and Sanchez-Ruiz, J. M. (1997) *Biochemistry 36*, 9616−9624.
41. Sosnick, T. R., Mayne, L., and Englander, W. (1996) *Proteins: Struct., Funct., Genet. 24*, 413−426.

BI980535T