Direct Kinetic Evidence for Folding via a Highly Compact, Misfolded State*

(Received for publication, June 4, 1999, and in revised form, July 7, 1999)

Maya J. Pandya‡§¶, Phil B. Williams‡, Christopher E. Dempsey‡, Peter R. Shewry§, and Anthony R. Clarke‡¶

From the ‡Molecular Recognition Centre and Department of Biochemistry, School of Medical Sciences, University of Bristol, University Walk, Bristol, BS8 1TD, United Kingdom and §Institute of Arable Crops Research, Long Ashton Research Station, Department of Agricultural Sciences, University of Bristol, Long Ashton, Bristol, BS18 9AF, United Kingdom

The 2 S seed storage protein, sunflower albumin 8 (SFA-8), contains an unusually high proportion of hydrophobic residues including 16 methionines (some of which may form a surface hydrophobic patch) in a disulfide cross-linked, α-helical structure. Circular dichroism and fluorescence spectroscopy show that SFA-8 is highly stable to denaturation by heating or chaotropic agents, the latter resulting in a reversible two-state unfolding transition. The small $m_F$ (−4.7 kcal mol$^{-1}$ at 10 °C) and $\Delta C_p$ (−0.95 kcal mol$^{-1}$ K$^{-1}$) values indicate that relatively little nonpolar surface of the protein is exposed during unfolding. Commensurate with the unusual distribution of hydrophobic residues, stopped-flow fluorescence data show that the folding pathway of SFA-8 is highly atypical, in that the initial product of the rapid collapse phase of folding is a compact nonnative state (or collection of nonnative states) that must unfold before acquiring the native conformation. The inhibited folding reaction of SFA-8, in which the misfolded state ($m_F = -0.95$ kcal mol$^{-1}$ at 10 °C) is more compact than the transition state for folding ($m_F = -2.5$ kcal mol$^{-1}$ at 10 °C), provides direct kinetic evidence for the transient misfolding of a protein.

The physical properties that direct a protein to its native fold are the subject of extensive experimental and theoretical investigation. It is generally accepted that the main driving forces in protein folding are formation of a distinct hydrophobic core by clustering of nonpolar groups (1) and local ordering of the backbone into elements of secondary structure according to the intrinsic properties of the amino acid residues and the need for pairing of backbone amide groups (2). The temporal development of these types of interaction, however, is more open to question. Hierarchic models of protein folding propose an initial formation of marginally stable local microdomains in which the secondary structure is well ordered, followed by their consolidation through the formation of longer sequence-range interactions and tertiary packing (3, 4). An alternative model suggests that global hydrophobic collapse drives an overall condensation of the polypeptide chain in the early stages of protein folding, which reduces the conformational possibilities and leads to the formation of secondary structure (5, 6).

Experimental studies of the nature of states arising in the folding reaction have led to starkly divergent conclusions. It is undeniable that collapsed intermediate states accumulate transiently during the refolding of most globular proteins, particularly those with polypeptide chains of less than 100 residues (7). These compact intermediates are formed rapidly (microsecond time scale) (8, 9) and contain extensive secondary structure, but they lack the fixed and near crystalline tertiary side-chain contacts characteristic of the native conformation. These intermediates may be viewed as productive “on-pathway” states that guide the protein to its native fold through organizing the backbone topology (7, 10) or as nonproductive “off-pathway” states, which are kinetically trapped because the rate-limiting energy barrier that divides them from the native state is raised (11, 12). In support of this latter view, lattice model simulations of heteropolymer organization suggest that partially folded intermediates contain stable, nonnative contacts that must be broken before the native structure is reached (13, 14).

Quantitative kinetic analyses of several proteins have been used to estimate the stability of such intermediates but do not determine whether these intermediates are productive or not (15–18). Late folding intermediates with nonnative interactions have been observed experimentally and can be explained by incorrect proline isomers in the case of ribonuclease A (19) and ribonuclease T1 (20), incorrect disulfide isomers in bovine pancreatic trypsin inhibitor (21, 22), and a misligated heme group in cytochrome c (16, 23). In addition, lysozyme folds by parallel pathways including a minor route without detectable intermediates (24–26). The highly stable intermediate formed on the major route of lysozyme folding is proposed to contain a nonnative aromatic cluster (27, 28). Folding of the intertwined dimeric DNA binding domain of the human papillomavirus E2 protein (29) and of the tryptophan repressor from Escherichia coli (30) is proposed to involve a nonnative, monomeric intermediate. Interestingly, the predominantly β-sheet protein, β-lactoglobulin, contains nonnative α-helical structure (31) in its early folding intermediate, while TEM-1 β-lactamase folds through a collapsed intermediate containing nonnative hydrophobic interactions (32).

In this paper, we examine the folding behavior of the 2 S seed storage protein, sunflower albumin 8 (SFA-8). 1 In preliminary

---

* This work received Cooperative Awards in Science and Engineering project funding from the Biotechnology and Biological Sciences Research Council, which also provided grant-aided support to Institute of Arable Crops Research. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

† To whom correspondence should be addressed: Dept. of Biochemistry, School of Biological Sciences, University of Sussex, Falmer, East Sussex, BN1 9QG, UK. Tel.: 44-1273-678683; Fax: 44-1273-678433.

‡ A Lister Institute Fellow.

---

1 The abbreviations used are: SFA-8, sunflower albumin 8; $m_F$, fraction of molecules in state $x$; $\Delta C_p$, heat capacity difference at constant pressure; $\Delta G_{x \rightarrow y}$, free energy difference between states $x$ and $y$; $\Delta G^*$, apparent activation free energy barrier; $\Delta H$, enthalpy difference; $\Delta S$,
experiments, this protein showed the unusual property of folding more rapidly in the presence of moderate concentrations of denaturant than it does in water. SFA-8 is known to form highly stable emulsions with oil/water mixtures (33), implying the presence of a large exposed surface of hydrophobic residues (34, 35). Proteins generally shield hydrophobic groups from aqueous solution (1). However, SFA-8 contains an unusually high proportion of hydrophobic residues including 16 methionines in a mature protein of 103 amino acids (36); we initially speculated that this may be the reason for the unorthodox folding behavior.

Using information from CD and NMR spectroscopy and sequence homology, we recently described a model for the structure of the protein based on the known structure of a related protein, a bifunctional a-amylase/trypsin inhibitor from seed of finger millet (38). The model structure consists of a bundle of four a-helices cross-linked by four disulfide bonds and supports the idea that the molecule has a large hydrophobic face. We describe here experiments that investigate the physical characteristics of the protein and probe its folding behavior. We find that the folding pathway is highly atypical, in that the initial product of the rapid collapse phase of folding is a highly compact state (or collection of nonnative states) that must unfold before passing over the rate-limiting transition barrier to form the native state.

**EXPERIMENTAL PROCEDURES**

**Purification of SFA-8**—Native SFA-8 was purified from sunflower seeds by cation exchange chromatography and reversed-phase high performance liquid chromatography, and the identity and integrity of the purified protein was confirmed by mass spectrometry. Freeze-dried Tatham, P. R. Shewry, and A. R. Clarke, submitted for publication.

**Secondary Structure Determination**—Far-UV CD spectra (185–250 nm) were recorded in 0.1-mm path length cells with a Jobin Yvon CD6 spectropolarimeter. A sample of native SFA-8 at 0.15 mM in 25 mM potassium phosphate buffer, pH 7, was examined at 25 °C and 90 °C. Data accumulations using 0.5-nm step resolution, 1-nm min−1 scan rate, and 2-nm bandwidth are reported with base-line subtraction. The fractional a-helical content (fH) of SFA-8 was estimated from the molar ellipticity measurement at 222 nm (θ192), using the algorithm (39),

\[ \theta_{222} = f_H \times (-42,500(1 - x/N)) + 640 + 640 \]  
(Eq. 1)

where x is the number of non-hydrogen-bonded peptide CO groups in the peptide, and N is the number of residues in the peptide chain.

Near-UV CD spectra (250–320 nm) of samples at 65 mM were recorded in a 0.5-cm path length cell at 25 and 90 °C. A sample of SFA-8 at 0.22 mM in 25 mM potassium phosphate buffer, pH 7, in the presence of 4 M GdmSCN was analyzed using a 0.1-mm path length cell at 25 °C.

**Fluorescence Spectroscopy**—Fluorescence emission spectra (300–500 nm) were recorded with a Perkin-Elmer LS-50B fluorescence spectro-photometer (excitation, 290 nm). Samples of SFA-8 at 0.25 mM in 25 mM potassium phosphate buffer, pH 7, were examined in the absence and presence of 4 M GdmSCN. Control samples of N-acetyltrypsinomamide (NATA) at 25 mM were used in order to quantify the fluorescence quenching effect of GdmSCN. Additional spectra were recorded over the pH range 2–10 in a 90 mM sodium borate/phosphate/acetate buffer. Stern-Volmer quenching constants (40) were measured by fitting the following equation using the corrected fluorescence intensities as follows,

\[ I = (I_0 - I_f) / I_0 - I_f \]  
(Eq. 3)

with temporary variables

\[ \alpha_F = K_{FU}(1 + K_{FU}) \]  
(Eq. 5)

\[ K_{FU} = K_{FU}^{(L)} \exp(m_1 x) \]  
(Eq. 6)

\[ \Delta G_{F(U)} = -RT \ln K_{FU} \]  
(Eq. 7)

where \( m_1 \) is the fraction of molecules in the folded state; \( K_{FU} \) is the equilibrium constant (FU) at a given solvent condition; \( K_{FU}^{(L)} \) is the equilibrium constant in water; and \( m_1 \) describes the sensitivity of the equilibrium between F and U to GdmSCN concentration (x) (see Fig. 1 in ref. 17). All data were fitted using the Marquardt nonlinear least-squares algorithm within the GraFit 3.00 analysis software. The change in free energy between F and U in water, \( \Delta G_{F(U)} \), was estimated using the equation,

\[ F = (F_0 - F_M)/(1 + c \cdot Q) + F_M \]  
(Eq. 2)

where \( F_0, F_M, \) and \( F_M \) are fluorescence intensities (measured, initial, and minimum, respectively), c is the Stern-Volmer constant, and Q is the concentration of iodide.

**NMR Spectroscopy**—NMRS spectra were obtained using a Jeol Alpha 500 MHz spectrometer. One-dimensional NMRS spectra were recorded for an amide deuterated sample of SFA-8 at 2 mM, which was titrated against increasing concentrations of deuterated GdmSCN (1.5, 2.45, 2.75, 3.5, and 4 mM), buffered by 50 mM disodium deuterium phosphate in 100% D2O, pH 6, at 27 °C. Two-dimensional double quantum-filtered correlation spectroscopy (42) of native SFA-8 and of denatured SFA-8 at 4 mM GdmSCN (in 20 mM disodium deuterium phosphate in 100% D2O, pH 6) was recorded at 27 °C. pH readings in D2O are direct measurements using a hydrogen electrode, without correction for the deuterium isotope effect. Spectra were acquired in phase-sensitive mode with quadrature detection and presaturation of the water signal using Dante pulse irradiation (43) during the relaxation decay. Two-dimensional spectra were obtained without spinning and with the irradiation and observation frequencies matched. The spectral width was set to ±3000 Hz around the residual solvent signal. Data sets were collected with 2048 real points in the t1 dimension and 512 t2 increments. These sets were zero-filled in t1, and shifted sine bell or sine squared bell window functions were applied in both dimensions before Fourier transformation and base-line correction. Data processing was performed using Felix 2.30 software (Bioyss Technologies Inc.).

**Equilibrium Denaturation**—Samples of SFA-8 at 2 μM and hen egg white lysozyme (Sigma, grade 1) at 10 μM were incubated with 0–6 mM guanidinium chloride (GdmCl, Sigma) or 0–5 mM GdmSCN in 50 mM potassium phosphate buffer, pH 7.2, for 16 h at 25 °C. Fluorescence emission was measured (excitation, 290 nm) for the equilibrated SFA-8 and lysozyme solutions at 340 and 380 nm, respectively. Further samples of SFA-8 were reduced using 1 mM DTT in 50 mM Tris/HCl, pH 7.2, containing 5 mM GdmSCN for 30 min. The reduced protein was diluted to 5 μM with 20 mM DTT in 50 mM Tris/HCl, pH 7.2 and then titrated against aliquots of 20 mM DTT containing 5 mM GdmSCN in the presence and absence of 0.4 mM Na2S2O3. The fluorescence emission at 340 nm was recorded after each addition (equilibrium was immediate).

The measured fluorescence values were plotted as a function of GdmSCN concentration. The effects of GdmSCN on the spectroscopic properties of the folded (F) and unfolded (U) states of the protein resulted in a slope of the base line at low and high GdmSCN concentrations, respectively; data were corrected for these effects. The fluorescence intensities relative to the unfolded state (I) were calculated from the following equation using the corrected fluorescence intensities as follows,

\[ I = (I_0 - I_f) / I_0 - I_f \]  
(Eq. 3)

with temporary variables

\[ \alpha_F = K_{FU}(1 + K_{FU}) \]  
(Eq. 5)

\[ K_{FU} = K_{FU}^{(L)} \exp(m_1 x) \]  
(Eq. 6)

\[ \Delta G_{F(U)} = -RT \ln K_{FU} \]  
(Eq. 7)

where R is the gas constant (1.9872 cal K−1 mol−1) and T is absolute
temperature (298 K).

Variation of Free Energy with Temperature—Further samples of SFA-8 at 5 mM were titrated against aliquots of 5 M GdmSCN at different temperatures (10–55 °C), and the fluorescence emission at 340 nm was recorded after each addition (equilibration was immediate). Values for the free energy change associated with transitions at each temperature were calculated as described for equilibrium denaturation experiments (with \( mU \) fixed at the mean value for all transitions). Data for the variation of the free energy change (\( \Delta G \)) with temperature (\( T \)) were fitted to the equation,

\[
\Delta G_T = \Delta H_{T_0} + \Delta C_p \cdot (T - T_0) + T \cdot \Delta S_{T_0} - \Delta C_p \cdot T \cdot \ln(T/T_0)
\]

(Eq. 8)

where \( \Delta H_{T_0} \) and \( \Delta S_{T_0} \) are the enthalpy and entropy changes, respectively, at an arbitrarily defined reference temperature (\( T_0 \)) and \( \Delta C_p \) is the change in heat capacity (at constant pressure) associated with the transition.

Denaturant Activity—Solutions of 0–5 M GdmSCN were prepared, to which saturating quantities of NATA were added. After equilibration for 24 h with shaking at 25 °C, absorbance readings were taken at 280 nm. Values for the free energy of transfer (\( \Delta G_S \)) of NATA from water to given concentrations of the denaturant GdmSCN were calculated from the equation,

\[
\Delta G_S = -RT \cdot \ln(A_{280}^w/A_{280}^x)
\]

(Eq. 9)

where \( A_{280}^w \) and \( A_{280}^x \) are the absorbance readings at 280 nm in water and \( x \) M denaturant, respectively. These values were plotted against denaturant concentration and fitted to the hyperbolic relationship,

\[
\Delta G_S = \Delta G_{SM} \cdot x/(C_{0.5} + x)
\]

(Eq. 10)

where \( \Delta G_{SM} \) is the maximum change in solvation energy at an infinite concentration of GdmSCN (\( x \)), and \( C_{0.5} \) is a denaturation constant that represents the concentration of denaturant required to reach half \( \Delta G_{SM} \).

Stopped-Flow Kinetics—Folding and unfolding rates were measured in a Hi-Tech SF-51 stopped-flow fluorescence spectrophotometer. An excitation wavelength of 298 nm was selected by a single monochromator from a mercury-xenon light source, and the fluorescence intensity above 320 nm was recorded using a cut-off filter. Each recorded transient was an average of at least five individual reactions. Folding reactions were initiated by mixing a 16 mM solution of SFA-8 containing 3 M GdmSCN in 50 mM Tris/HCl, pH 7.2, with either 2 or 5 volumes of buffer or buffered GdmSCN to yield the required final GdmSCN concentration. Unfolding reactions were initiated by mixing a 16 mM solution of SFA-8 containing 2 M GdmSCN in 50 mM Tris/HCl, pH 7.2, with either 2 or 5 volumes of buffered GdmSCN. All transients were described by single-exponential functions.

Folding and unfolding reactions were recorded at 5, 10, 15, and 20 °C. At 10 °C, folding reactions were recorded as controls for the effects of aggregation (using 80 mM SFA-8), salt dependence, and disulfide isomerization of the unfolded protein at high pH. Salt effects were investigated by mixing 16 mM SFA-8 in 3 M GdmSCN against 2 volumes of buffered NaCl to yield the required final NaCl concentration at 1 M GdmSCN. Disulfide isomerization was prevented by mixing 16 mM SFA-8 in 3 M GdmSCN in 50 mM disodium phosphate buffer, pH 5, against 2 volumes of buffer or buffered GdmSCN, pH 8, to yield the required final GdmSCN concentration at pH 7.2. In addition, 15 mM NATA in 3 M GdmSCN was mixed with 2 or 5 volumes of buffer to demonstrate that the fluorescence data obtained for SFA-8 are not an artifact of the mixing of solutions (on the appropriate time scale).

Data from the folding and unfolding reactions were modeled by the three-state system,
where $M$ represents the misfolded state, $K_{MU}$ is the equilibrium constant ($M/U$), and $k_{U\rightarrow F}$ and $k_{F\rightarrow U}$ are the folding and unfolding rate constants, respectively. Natural logarithms of the observed rate constants ($k$) for the folding and unfolding reactions were plotted as a function of GdmSCN concentration and fitted to the linear free energy relationship (17),

$$k = k_{F\rightarrow U} + k_{U\rightarrow F}(1 + K_{MU})$$

(Eq. 11)

with temporary variables

$$k_{F\rightarrow U} = k_{F\rightarrow U}(W) \exp(-m_T \cdot x)$$

(Eq. 12)

$$k_{U\rightarrow F} = k_{U\rightarrow F}(W) \exp(m_U - m_T) \cdot x$$

(Eq. 13)

$$K_{MU} = K_{MU}(W) \exp(m_U - m_M) \cdot x$$

(Eq. 14)

The subscript $W$ denotes constants in water; $m_T$, $m_U$, and $m_M$ are the linear free energy slopes for the unfolded, transition, and misfolded states, respectively, in the process relative to the folded state; $x$ is the GdmSCN concentration. These data were fitted using proportional weighting within the GraFit analysis software. The equilibrium constant ($F/U$) in water, $K_{F/U(W)}$, was estimated using the following equation.

$$K_{F/U(W)} = k_{U\rightarrow F}(W)/k_{F\rightarrow U}(W)$$

(Eq. 15)

The apparent activation energy of the transition state, $\Delta G^1_{U\rightarrow F(W)}$, was estimated using the equation,

$$\Delta G^1_{U\rightarrow F(W)} = -RT \cdot \ln(k_{U\rightarrow F(W)/h})$$

(Eq. 16)

by arbitrarily assuming a barrierless interconversion rate $h'$ of $10^{10}$ s$^{-1}$, the jump time for a water molecule (44).

Fluorescence amplitudes ($I_x$) in the folding reaction of SFA-8 were fitted to the equation,

$$I_x = I_U \cdot (a_U + a_M - 1) + I_F \cdot a_U - I_F \cdot a_M$$

(Eq. 17)

with temporary variables

$$a_U = 1/(1 + K_{FU})$$

(Eq. 18)

$$a_F = K_{FU}/(1 + K_{FU})$$

(Eq. 19)

$$a_M = K_{MU}(1 + K_{MU})$$

(Eq. 20)

where $I_U$, $I_F$, and $I_M$ are the fluorescence intensities (unfolded, folded, and misfolded, respectively) and $a_U$, $a_F$, and $a_M$ are the proportion of molecules unfolded, folded, and misfolded. Amplitudes in the unfolding reaction were fitted to the following equation.

$$I_x = I_F \cdot (1 - a_F) - I_U \cdot a_U$$

(Eq. 21)

RESULTS

Stability—SFA-8 is a $2 S$ seed storage protein of 103 amino acids with four disulfide bonds and a relatively large proportion of hydrophobic residues (16 Met, nine Leu, three Ile, three Tyr, two Val, one Trp) (36). A structural model based on circular dichroism, nuclear magnetic resonance, and sequence homology data comprises a disulfide cross-linked, antiparallel bundle of four $\alpha$-helices with a highly hydrophobic face.

Far-UV (amide) CD spectroscopy of the native protein at 25 °C (Fig. 1A) showed a maximum at 190 nm and minima at 209 and 221 nm, which are characteristic of a protein with a high content of $\alpha$-helical structure. The $\alpha$-helical content, estimated to be about 30% from the signal at 222 nm (39), was highly stable to heating to 90 °C (Fig. 1A). Comparison of near-UV (aromatic) CD spectra (Fig. 1B), which detect aromatic side chains in asymmetric environments, indicates that the tertiary structure of SFA-8 is also fairly resistant to heating at 90 °C.

Unfolding of SFA-8 was only achieved by using a high concentration of a powerful denaturing agent 4 M GdmSCN (45). The activity of this denaturant (46) was demonstrated to be approximately linear to 3.5 M GdmSCN with a denaturation constant ($C_{MU}$) determined as 6.5 mM (described under “Experimental Procedures”). Fluorescence emission spectra of SFA-8 in the native state and after chemical denaturation are presented in Fig. 2. Spectra of native SFA-8 did not vary over the pH range 2–10 (data not shown). Denaturation of SFA-8 was associated with a decrease in fluorescence intensity together with a red shift in emission maximum, caused by the exposure of its buried single tryptophan residue (Trp$^{76}$) to a more polar environment upon unfolding. Quenching of intrinsic fluorescence with sodium iodide was measured for SFA-8 in the native state and after reduction and S-carboxymethylation (to prevent reoxidation) of the disulfides. Compared with NATA, which has a Stern-Volmer constant of 10.5 (± 0.06) M$^{-1}$ in these conditions, native SFA-8 has a constant of 6.0 (± 0.06) M$^{-1}$, indicative of a buried tryptophan, whereas the reduced and S-carboxymethylated protein demonstrated higher solvent accessibility with a constant of 5.0 (± 0.1) M$^{-1}$.

One-dimensional NMR spectra of denatured SFA-8 at increasing concentrations of GdmSCN are shown in Fig. 3. At 1.5 M GdmSCN, the protein is clearly folded with well dispersed signals including ring current-shifted methyl protons (0–0.5 ppm), methionine methyl singlets (2 ppm), and exchange-stable amide protons. Increasing the denaturant concentration caused the loss of upfield-shifted methyl protons due to loss of...
Folding via a Highly Compact, Misfolded State

Unfolding of the protein is indicated by the loss of dispersion of signals, collapse of methyl singlets, and loss of exchange-stable amides. This transition was complete at 3.5 M GdmSCN and clearly demonstrates that SFA-8 is completely unfolded in the presence of 4 M GdmSCN. Further evidence comes from the aCH-/bCH region of a double quantum-filtered correlation spectroscopy spectrum of SFA-8 in 4 M GdmSCN (data not shown), which displays no spectral dispersion and is highly similar to a spectrum simulated using chemical shift values obtained for amino acid residues in random coil conformation (47).

Folding and Unfolding Reactions—CD spectroscopy at temperatures up to 90 °C demonstrated that SFA-8 is highly thermostable. Chemical denaturation and stopped-flow kinetics were used to characterize the stability of SFA-8 and the dynamics of the folding/unfolding process. Changes in fluorescence intensity for SFA-8 were monitored as a function of denaturant concentration using either GdmCl or GdmSCN. SFA-8 was not unfolded at 6 M GdmCl, but unfolding did occur in the presence of the stronger denaturant GdmSCN, with a midpoint of 2.6 M (Fig. 4A). Hen egg white lysozyme, a protein of similar size (129 amino acids) with four disulfide bonds that is frequently used as a model in protein folding studies, was unfolded over a narrower concentration range with a midpoint of 1.4 M in GdmSCN (Fig. 4A). Chemical denaturation of both proteins in GdmSCN was reversible, giving unfolding transitions consistent with cooperative two-state unfolding mechanisms at equilibrium. When the reduced form of SFA-8 was titrated with GdmSCN, the fluorescence decreased as a monotonic curve but lacked a sigmoidal transition as seen in the oxidized state. The addition of the cosmotropic agent sodium sulfate to the reduced form led to an increase in the fluorescence of the species, suggesting a greater level of compactness.

The slope of an unfolding transition (the m_W value) depends on the change in solvation of hydrocarbon between the folded (F) and unfolded (U) states (48, 49). A small m_W value of \(-4.6 \pm 0.1\) M\(^{-1}\) was determined for SFA-8 at 25 °C, compared with a value of \(-10.5 \pm 0.4\) M\(^{-1}\) for lysozyme. Therefore, SFA-8 exposes relatively little new nonpolar surface during the unfolding reaction. The change in free energy between the folded and unfolded states in the absence of denaturant, \(\Delta G_{U \rightarrow F, pW}\), was extrapolated from data for the unfolding of SFA-8 by denaturation with GdmSCN as \(-7.0 \pm 0.5\) kcal mol\(^{-1}\) at 25 °C (Table 1).

Values for the free energy change associated with unfolding transitions at different temperatures (10–55 °C) were calculated from equilibrium denaturation data and plotted as a function of temperature (Fig. 4B). The transition midpoint barely varies with temperature indicating a small change in heat capacity (\(\Delta C_p\)) during unfolding. The data were fitted to yield thermodynamic constants for the enthalpy change (\(\Delta H_{U \rightarrow F, pW}\)) and the entropy change (\(\Delta S_{U \rightarrow F, pW}\)). The heat capacity change (\(\Delta C_p\)) was determined for SFA-8 as \(-7.0 \pm 0.5\) kcal mol\(^{-1}\) at 25 °C (Table 1).

Values at different temperatures (10–55 °C) were calculated from equilibrium denaturation data and plotted as a function of temperature. The transition midpoint barely varies with temperature indicating a small change in heat capacity (\(\Delta C_p\)) during unfolding. The data were fitted to yield thermodynamic constants for the enthalpy change (\(\Delta H_{U \rightarrow F, pW}\)) and the entropy change (\(\Delta S_{U \rightarrow F, pW}\)). The heat capacity change (\(\Delta C_p\)) was determined for SFA-8 as \(-7.0 \pm 0.5\) kcal mol\(^{-1}\) at 25 °C (Table 1).
Hence, there is a smaller entropic penalty upon folding allowing the desolvation effect to dominate.

The change in heat capacity between the folded and unfolded states of a protein, \( \Delta C_p \), is dependent on the amount of surface exposed to solvent upon unfolding and typically has an average value of 14.2 cal mol\(^{-1}\) K\(^{-1}\) residue\(^{-1}\) (49). Considering the small magnitude of \( \Delta C_p \), it is not surprising that the \( \Delta C_p \) value obtained for SFA-8 is unusually small (Table I) and suggests unusually rapid folding/unfolding dynamics. As expected, the expression for the virtual folding rate in the absence of denaturant \( k_U^{\text{obs}} \) (see “Experimental Procedures”),

\[
h_{\text{obs}} = h_F^{\text{obs}} + h_{\text{U-F}}/(1 + K_{M,U})
\]

which describes a three-state mechanism (\( M \) to \( U \) to \( F \)) in which the off-pathway misfolded state (\( M \)) rapidly equilibrates with the unfolded state (\( U \)). Importantly, the kinetic parameters (Table I) obtained from the rate profile are consistent with the value of \( K_{M,U} \) derived from the equilibrium data.

The broken line a (Fig. 5B) is an extrapolation of the unfolding limb with a slope that defines \( -m_T \) and a y axis intercept that defines the unfolding rate in water \( k_{U-F}^{\text{obs}} \) (0.11 ± 0.03 s\(^{-1}\)). The folding reaction near the cusp of the rate profile is effectively two-state because the misfolded state is not significantly populated, such that the expression \( k_{U-F}^{\text{obs}}/(1 + K_{M,U}) \) reduces to \( k_{U-F}^{\text{obs}} \). Therefore, extrapolation of line b (Fig. 5B) gives a slope that defines \( m_T^{\text{E}} - m_T^{\text{F}} \) and an intercept that defines the virtual folding rate in the absence of denaturant and in the absence of a misfolded state, \( k_{U-F}^{\text{obs}}/(1.7 \times 10^4 \pm 5 \times 10^3 \text{ s}^{-1}) \). When the misfolded state is populated, the expression \( k_{U-F}^{\text{obs}}/(1 + K_{M,U}) \) reduces to \( k_{U-F}^{\text{obs}}/K_{M,U} \). Therefore, the unusual positive slope of the folding limb (region c in Fig. 5B) defines \( (m_M - m_T) \) and its intercept defines \( k_{U-F}^{\text{obs}}/K_{M,U} \) showing that the misfolded state is more compact (less negative \( m \) value) than the transition state. The value for \( m_T^{\text{E}} (-4.7 \pm 0.1 \text{ M}^{-1}) \) is essentially the same as that determined by equilibrium denaturation (Table I). The values of \( m_{U-F} \), \( m_T \) (0.95 ± 0.09 M\(^{-1}\)), and \( K_{M,U} \) (1.2 × 10^5 ± 3.6 × 10^5 M\(^{-1}\)) were used to produce an equilibrium profile of the proportion of molecules in the misfolded state rather than the unfolded state of SFA-8 (\( \alpha \)) in Fig. 4A.

The free energy difference between the unfolded and misfolded states (\( \Delta G_{U-M}^{\text{obs}} = 0.4 \pm 0.1 \text{ kcal mol}^{-1} \)) and between the unfolded and transition states (\( \Delta G_{U-T}^{\text{obs}} = 3.7 \pm 0.1 \text{ kcal mol}^{-1} \), assuming a barrierless interconversion rate of 10^18 s\(^{-1}\)) was estimated from the kinetic parameters. As expected from the positive slope \( (m_U - m_T) \), the misfolded state \( (m_T = -0.95 \pm 0.09 \text{ M}^{-1}) \) is more compact than the transition state \( (m_T = -2.5 \pm 0.1 \text{ M}^{-1}) \). The degree of solvation or compactness \( (m_M/m_U) \) of each state (relative to the folded state) may be used as a reaction coordinate when plotted against the free energy of each state (relative to the folded state) as shown in Fig. 6A. The relationship between denaturant concentration and the free energy of the folded, misfolded, transition, and unfolded states

\[
\begin{array}{cccc}
\text{Equilibrium parameter} & \text{Value} & \text{Kinetic parameter} & \text{Value} \\
\hline
m_U & -4.6 \pm 0.1 \text{ M}^{-1} & m_U & -4.7 \pm 0.1 \text{ M}^{-1} \\
K_{U/F} & 1.5 \times 10^3 \pm 1.6 \times 10^4 & m_T & -2.5 \pm 0.1 \text{ M}^{-1} \\
\Delta G_{U-F} & -6.7 \pm 0.1 \text{ kcal mol}^{-1} & m_M & -0.95 \pm 0.09 \text{ M}^{-1} \\
\Delta G_{U-T} & -0.95 \pm 0.04 \text{ kcal mol}^{-1} & k_{U-F}^{\text{obs}} & 1.7 \times 10^4 \pm 5 \times 10^3 \text{ s}^{-1} \\
\Delta H_{U-F} & 15.8 \pm 0.8 \text{ kcal mol}^{-1} & k_{U-F}^{\text{obs}} & 0.11 \pm 0.03 \text{ s}^{-1} \\
\Delta S_{U-F} & +8.2 \times 10^{-3} \pm 2.8 \times 10^{-3} \text{ kcal mol}^{-1} \text{ K}^{-1} & K_{M,U} & 1.2 \times 10^5 \pm 3.6 \times 10^5 \\
\hline
\text{At 25 °C} & & \Delta G_{U-F} & +7.5 \pm 0.1 \text{ kcal mol}^{-1} \\
\Delta G_{U-T} & +0.95 \pm 0.04 \text{ kcal mol}^{-1} \text{ K}^{-1} & \Delta G_{U-M} & -4.0 \pm 0.1 \text{ kcal mol}^{-1} \\
\Delta H_{U-T} & +1.5 \pm 0.3 \text{ kcal mol}^{-1} & & \\
\Delta S_{U-T} & +3.3 \times 10^{-3} \pm 1.1 \times 10^{-3} \text{ kcal mol}^{-1} \text{ K}^{-1} & &
\end{array}
\]
(with the folded protein as the reference state) is shown in Fig. 6B. As the relative free energy value for each of the states ($F$, $M$, $T$, and $U$) varies with denaturant concentration (proportional to exposed nonpolar surface), the rates of interconversion between these states will alter, resulting in different observed rate constants and different final equilibrium populations. The point at which the slopes for two states intersect corresponds to the midpoint of the appropriate equilibrium transition in Fig. 4A.

The effect of denaturant concentration on the population of the kinetic states at equilibrium (Fig. 7A) demonstrates that the misfolded state is not populated during the equilibrium
DISCUSSION

Stability—The disulfide cross-linked, α-helical protein, sunflower albumin 8 (SFA-8) shows high stability to heating, since the CD spectrum is relatively unaffected by heating to 90 °C. Unfolding of native SFA-8 was only achieved in the presence of high concentrations of a strong chaotrope agent. One- and two-dimensional NMR experiments demonstrate that SFA-8 is completely unfolded in 4 M GdmSCN. Fluorescence data show that it exhibits a cooperative, two-state unfolding transition in the oxidized form. However, the reduced protein is completely unfolded at a lower denaturant concentration in a noncooperative transition. Unsurprisingly, we conclude that the disulfide bonds are important in maintaining the stability of the protein’s native fold.

Equilibrium unfolding of oxidized SFA-8 has a transition midpoint that is consistent with the cusp of the rate profile, confirming that the folding and unfolding reactions are fully reversible. The sensitivity of the equilibrium to denaturant concentration, the $m_U$ value, is a qualitative measure of the degree to which nonpolar (i.e., hydrocarbon) groups in the protein are newly exposed to solvent upon unfolding (46, 51, 52). Lysozyme, a protein of similar size to SFA-8 which also contains four disulfide bonds, unfolds with a $m_U$ of $-10.5$ M$^{-1}$ using GdmSCN as the denaturant. SFA-8 unfolded with a shallower transition and hence a smaller $m_U$ value ($-4.6$ M$^{-1}$). Therefore, SFA-8 buries less hydrocarbon upon folding and is less sensitive to denaturant concentration.

SFA-8 is unusually stable to solvent denaturation with an unfolding midpoint of 2.6 M GdmSCN at 25 °C, as compared with 1.4 M GdmSCN for lysozyme. However, the free energy change for folding of SFA-8 ($\Delta G_{U\rightarrow P\rightarrow W}$) is $-7.0$ kcal mol$^{-1}$ at 25 °C is smaller than for lysozyme; hence, the high midpoint is due to a smaller change in solvation of hydrocarbon rather than a more negative $\Delta G_{U\rightarrow P\rightarrow W}$. The high stability of SFA-8 to heating is due to an unusually small heat capacity change ($\Delta C_p(U\rightarrow P\rightarrow W) = -0.95 \pm 0.04$ kcal mol$^{-1}$ K$^{-1}$), which also depends on the amount of surface exposed to solvent upon unfolding. This small change in exposure of hydrocarbon, together with the high proportion of hydrophobic residues present in SFA-8, implies that the folded protein contains an unusually large amount of exposed hydrocarbon and correspondingly less buried hydrocarbon.

Folding Pathway—The most unusual feature of SFA-8 is its folding kinetics. The kinetic profile for SFA-8 shows deviation from linearity at low denaturant concentrations (below 2 M GdmSCN). This represents an inhibited folding reaction in which an off-pathway misfolded state becomes populated and must unfold before the native state is attained, so that an unfolding step is part of the folding pathway. The true folding rate constant ($k_{U\rightarrow P\rightarrow W}$) was determined as $1.7 \times 10^4 \pm 5 \times 10^3$ s$^{-1}$ at 10 °C, which implies a time constant of just 60 μs, and this of course would be faster at elevated temperatures. This time constant is not much smaller than the values of 130 and 300 μs determined for folding of reduced cytochrome c at 40 °C (53) and the 80-residue domain of the λ repressor at 37 °C (54), respectively; 1 μs is proposed as the shortest time in which a small protein can fold (9). In addition, the kinetic parameters for SFA-8 are consistent with a linear relationship that has been observed between the unfolding rate constants and the position of the transition states along the folding pathways of a large number of proteins (55).

The misfolded state of SFA-8 (M) has a fractional solvent exposure ($m_M/m_U$) of 0.2 and is clearly more compact than the transition state between the unfolded and folded states ($m_L/m_U = 0.53$). The fluorescence intensity of the misfolded state ($I_M = 0.84$) is closer to that of the folded state ($I_U = 1$).
mally, proteins have a pattern of hydrophobic residues, which form a well-defined core during the initial hydrophobic collapse. This course pattern of nonpolar residues drives rapid formation of a topologically native intermediate. In SFA-8, the definition of core and surface, encoded by the pattern of residues along the chain, is less clear. Consequently, nonpolar residues destined for the surface in the folded state are likely to be driven out of the solvent and into a nonnative core prior to the rate-limiting acquisition of precise side-chain interactions.

This is not the first evidence for nonnative structure in folding intermediates. In the dead time intermediate of β-lactamase folding, indole groups are shown to be more buried than in the native state (32), and nonnative α-helical structure has been seen in the early folding intermediate of β-lactoglobulin (31). Such observations can be argued to reflect the balance between the influence of intrinsic secondary structure preferences in the first stage of folding and tertiary interactions in determining the final folded structure (31, 37). However, SFA-8 represents an extreme and thus far unprecedented case of misfolding in that the initially collapsed ensemble has to unfold extensively before passing through a less compact transition state. This behavior is manifest in the real increase in the folding rate of SFA-8 in conditions that suppress hydrophobic interactions, making this protein an exception to a general rule.

Acknowledgments—We are grateful to M. J. Parker, R. B. Sessions, and C. B. Clarke for critical assessment of the manuscript.

REFERENCES

1. Dill, K. A. (1990) Biochemistry 29, 7133–7155
2. Dyson, H. J., and Wright, P. E. (1995) Curr. Opin. Struct. Biol. 5, 60–65
3. Baldwin, R. L. (1989) Trends Biochem. Sci. 14, 291–294
4. Karplus, M., and Weaver, D. L. (1994) Protein Sci. 3, 650–668
5. Chan, H. S., and Dill, K. A. (1993) Phys. Today 46, 24–32
6. Pitsyn, O. B. (1995) Curr. Opin. Struct. Biol. 5, 74–78
7. Roder, H., and Colom, W. (1997) Curr. Opin. Struct. Biol. 7, 15–28
8. Ballew, R. M., Sabelko, J., and Gruebele, M. (1996) Proc. Natl. Acad. Sci. U. S. A. 93, 5759–5764
9. Eaton, W. A., Muñoz, V., Thompson, P. A., Chan, C.-K., and Hofrichter, J. (1997) Curr. Opin. Struct. Biol. 7, 10–14
10. Parker, M. J., Dempsey, C. E., Lorch, M., and Clarke, A. R. (1997) Biochemistry 36, 13396–13405
11. Creighton, T. E., Darby, N. J., and Kemmink, J. (1996) FASEB J. 10, 110–118
12. Schindler, T., Herrler, M., Marahiel, M. A., and Schmid, F. X. (1995) Nat. Struct. Biol. 2, 663–672
13. Bryngelson, J. D., Onuchic, J. N., Socci, N. D., and Wolynes, P. G. (1995) Proteins Struct. Funct. Genet. 21, 167–195
14. Thiramaiyal, D., and Guo, Z. (1995) Biopolymers 35, 137–140
15. Matsouche, A., Kellis, Jr., J. T., Serrano, L., Bycroft, M., and Fersht, A. R. (1990) Nature 346, 440–445
16. Sosnick, T. R., Mayne, L., Hiller, R., and Englander, S. W. (1994) Nat. Struct. Biol. 1, 149–156
17. Parker, M. J., Spencer, J., and Clarke, A. R. (1995) J. Mol. Biol. 253, 771–786
18. Khorasanizadeh, S., Peters, I. D., and Roder, H. (1996) Nat. Struct. Biol. 3, 193–205
19. Houri, W. A., and Scheraga, H. A. (1996) Biochemistry 35, 11719–11733
20. Balbáth, J., Steegborn, C., Schindler, T., and Schmid, F. X. (1999) J. Mol. Biol. 295, 829–842
21. Weissman, J. S., and Kim, P. S. (1992) Proc. Natl. Acad. Sci. U. S. A. 89, 9900–9904
22. Zhang, J.-X., and Goldenberg, D. P. (1993) Biochemistry 32, 14075–14081
23. Stuck, T. R., Mayne, L., and Englander, S. W. (1996) Proteins Struct. Funct. Genet. 24, 413–426
24. Miranker, A., Robinson, C. V., Radford, S. E., Aplin, R. T., and Dobson, C. M. (1993) Science 262, 896–900
25. Miranker, A., Robinson, C. V., Radford, S. E., and Dobson, C. M. (1996) FASEB J. 10, 93–101
26. Wobbege, G., and Königshofer, T. (1997) J. Mol. Biol. 270, 294–304
27. Rothwarf, D. M., and Scheraga, H. A. (1995) Biochemistry 34, 13797–13807
28. Kazmirska, S. L., and Daggett, V. (1996) J. Mol. Biol. 264, 793–806
29. Mok, Y.-K., Bycroft, M., and de Prat-Gay, G. (1996) Nat. Struct. Biol. 3, 711–717
30. Shao, X., and Matthews, C. R. (1998) Biochemistry 37, 7850–7858
31. Araki, M., Ikura, T., Semisotnov, G. V., Khara, H., Amemiya, Y., and Kuwajima, K. (1998) J. Mol. Biol. 275, 149–162
32. Vanhove, M., Lejeune, A., Guillaume, G., Virden, R., Pain, R. H., Schmid, F. X., and Frère, J.-M. (1998) Biochemistry 37, 1941–1950
33. Guesjen, J., Popineau, Y., Anisimova, I. N., Fido, R. J., Shewry, P. R., and Tatham, A. S. (1996) J. Agric. Food Chem. 44, 1184–1189
34. Fligner, K. L., and Mangino, M. E. (1991) ACS Symp. Ser. 454, 1–12
35. Li-Chan, E., and Nakai, S. (1991) ACS Symp. Ser. 448, 193–212
36. Kort, A. A., Caldwell, J. B., Lilley, G. G., and Higgins, T. J. V. (1991) Eur.
References

37. Minor, Jr., D. L., and Kim, P. S. (1996) Nature 380, 730–734
38. Strobl, S., Mühlhahn, P., Bernstein, R., Wütscheck, R., Maskos, K., Wunderlich, M., Huber, R., Glockshuber, R., and Holak, T. A. (1995) Biochemistry 34, 8281–8293
39. Rohr, C. A., Chakrabartty, A., and Baldwin, R. L. (1996) Protein Sci 5, 2623–2637
40. Weller, A. (1961) Prog. React. Kinet. 1, 187–214
41. Imoto, T., and Yamada, H. (1989) in Protein Function: A Practical Approach (Creighton, T. E., ed) pp. 247–279, IRL Press, Oxford
42. Rance, M., Sørensen, O. W., Bodenhausen, G., Wagner, G., Ernst, R. R., and Wuthrich, K. (1983) Biochem. Biophys. Commun. 117, 479–485
43. Morris, G. A., and Freeman, R. (1978) J. Magn. Reson. 29, 433–462
44. Atkins, P. W. (1995) Physical Chemistry, p. 855, Oxford University Press, Oxford
45. Baldwin, R. L. (1996) Biophys. J. 71, 2056–2063
46. Staniforth, R. A., Burston, S. G., Smith, C. J., Jackson, G. S., Badcoe, I. G., Atkinson, T., Hollbrook, J. J., and Clarke, A. R. (1993) Biochemistry 32, 3842–3851
47. Wishart, D. S., Bigam, C. G., Holm, A., Hodges, R. S., and Sykes, B. D. (1995) J. Biomol. NMR 5, 67–81
48. Tanford, C. (1970) Adv. Protein Chem. 24, 1–95
49. Myers, J. K., Pace, C. N., and Scholtz, J. M. (1995) Protein Sci. 4, 2138–2148
50. Privalov, P. L., and Gill, S. J. (1988) Adv. Protein Chem. 39, 191–224
51. Pace, C. N. (1975) CRC Crit. Rev. Biochem. 3, 1–43
52. Shortle, D., Meeker, A. K., and Freire, E. (1988) Biochemistry 27, 4761–4768
53. Pascher, T., Chesick, J. P., Winkler, J. R., and Gray, H. B. (1996) Science 271, 1558–1560
54. Huang, G. S., and Oas, T. G. (1995) Proc. Natl. Acad. Sci. U. S. A. 92, 6878–6882
55. Parker, M. J., Lorch, M., and Clarke, A. R. (1998) Biochemistry 37, 2538–2545