Unusual Effects of an Engineered Disulfide on Global and Local Protein Stability†

Stephen F. Betz,‡,§ Jennifer L. Marmorino,‡ Aleister J. Saunders,|| Donald F. Doyle,‡ Gregory B. Young,|| and Gary J. Pielak*,‡

Department of Chemistry and Department of Biochemistry & Biophysics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-3290

Received December 4, 1995; Revised Manuscript Received March 14, 1996⊗

ABSTRACT: The global and local stabilities of a eukaryotic ferricytochrome *c* variant with an engineered disulfide are examined. The disulfide connects position 20, which is usually a valine, to position 102, which is usually a threonine. The cross-linked variant is ∼1.2 kcal mol^{-1} *less* stable than the wild-type protein at 298 K, pH 4.6, in H_2O and D_2O. Circular dichroism studies show that the decreased stability results from structure-induced stabilization of the denatured state [Betz, S. F., & Pielak, G. J. (1992) *Biochemistry 31*, 12337−12344]. Here, we use proton chemical shift, paramagnetic shift, and amide proton exchange data to obtain atomic level structural and energetic information. Chemical and paramagnetic shift data indicate only minor native state structural changes. Local stability is obtained from amide proton−deuterium exchange data, using model peptide intrinsic exchange rates. As expected, the exchange data indicate that cross-link incorporation decreases the majority of local stabilities. Near the cross-link, however, local stability seems to *increase* despite the overall global stability decrease. Furthermore, local stability changes for hydrophobic core residues seem to be *greater* than the global stability change. We interpret these observations as cross-link-induced changes in exchange competent states and relate them to changes in the denatured state.

The ability to control protein stability remains an elusive goal. An important step toward attaining this goal will come from knowing what stabilizes native states and from exploring the energetics of intermediate and denatured states. A common approach to increasing stability is the introduction of disulfides (Betz, 1993b).

Recently, we examined the structure and stability of a *Saccharomyces cerevisiae* iso-1-cytochrome *c* variant with an engineered intramolecular disulfide using calorimetry, circular dichroism (CD), and absorbance spectroscopy (Betz & Pielak, 1992). We compared these data to those for the C102T variant, in which the cysteine at position 102[1] is replaced by threonine (Cutler et al., 1987). The C102T variant is structurally identical to the authentic wild-type protein (Gao et al., 1990, 1991; Berghuis & Brayer, 1992), but it is more amenable to biophysical studies because the natural protein forms intermolecular disulfide-bonded dimers (Bryant et al., 1985; Betz & Pielak, 1992). The V20C variant, which is modeled after bullfrog cytochrome *c* (Brems et al., 1982), possesses an intramolecular disulfide between the natural cysteine at position 102 and an engineered cysteine at position 20. The disulfide connects the C-terminal helix (residues 87−102) to the sequence following the N-terminal helix (residues 2−14) and the heme-binding region (residues 14−18). Perpendicular packing of these helices is the most conserved element of cytochrome *c* tertiary structure (Mathews, 1985). Hereafter we refer to the C102T variant as the wild-type (wt) protein and the V20C variant as the cross-linked variant.

The cross-linked variant is *less stable* than the wt protein at 298 K. This stability decrease results from structure-induced stabilization of the denatured state (Betz & Pielak, 1992). This decrease is opposite to what is expected for incorporation of a disulfide that does not affect the native state (Flory, 1956; Betz, 1993b; Harrison & Sternberg, 1994). Here, we compare the structures, global stabilities, and local stabilities of the ferri-forms of the cross-linked variant and the wt protein.

To compare structures we analyze paramagnetic shifts. Cytochromes *c* exist in two oxidation states. In the ferro [Fe(II) or reduced]-form, the heme iron is diamagnetic (low spin, d^6). In the ferri [Fe(III) or oxidized]-form, the heme iron is paramagnetic with one unpaired electron (low spin, d^5). The paramagnetic shift, δ_{ps}, is the chemical shift of a particular proton in the oxidized state minus the chemical shift of that same proton in the reduced state. Comparison of δ_{ps} between variant proteins allows examination of structural changes (Gao et al., 1991; Auld et al., 1993). Paramagnetic shift comparisons are a more sensitive indicator of structural changes than even nuclear Overhauser effect comparisons (Gochin & Roder, 1995).

To compare global stabilities we examine thermal denaturation of the ferri-forms in H_2O and D_2O using CD. Global stability is ΔG_D, the free energy of denaturation. Both proteins exhibit reversible, two-state denaturation (Betz & Pielak, 1992; Cohen & Pielak, 1994).

To compare local stabilities we monitor hydrogen−deuterium (HD) exchange using NMR (Berger & Linderstrøm-Lang, 1957; Hvidt & Nielsen, 1966; Englander & Kallenbach, 1984; Baldwin, 1993). The observed HD

† Supported by donors of the Petroleum Research Fund, administered by the American Chemical Society (21597-G), and the NIH (GM42501).

* Corresponding author. Tel: (919) 966-3671. FAX: (919) 962-2388. E-mail: gary_pielak@unc.edu.

‡ Department of Chemistry.

§ Current address: Department of Biochemistry and Biophysics, Johnson Research Foundation, University of Pennsylvania, Philadelphia, PA 19104.

|| Department of Biochemistry & Biophysics.

⊗ Abstract published in *Advance ACS Abstracts,* May 15, 1996.

[1] The vertebrate cytochrome *c* numbering system is used (Moore & Pettigrew, 1990).

S0006-2960(95)02855-8 CCC: $12.00

exchange rate constant, k_{obs}, depends on several features, including the solvent accessibility of the proton and sequence specific effects (Molday et al., 1972; Bai et al., 1993). The most common interpretation of backbone amide HD exchange is the local unfolding model, eq 1. In this model,

$$N(H) \underset{k_2}{\overset{k_1}{\rightleftharpoons}} OP(H) \xrightarrow{k_{INT}} OP(D) \underset{k_1}{\overset{k_2}{\rightleftharpoons}} N(D) \quad (1)$$

the native state, N, is in equilibrium with exchange competent open states, OP, which arise from normal thermal fluctuations. In these high-energy states amide protons exchange with deuterons.

Under conditions that favor the native state, the rate constant for closing, k_2, is much greater than the rate constant for opening, k_1, so that

$$k_{obs} = \frac{k_1 k_{INT}}{(k_2 + k_{INT})} \quad (2)$$

where k_{INT} is the intrinsic rate of HD exchange from the open state. Structural opening, k_1, determines the observed rate constant k_{obs}, if $k_2 \ll k_{INT}$, whereas the ratio k_{INT}/k_2 determines k_{obs} if $k_2 \gg k_{INT}$. The former is the EX_1 mechanism, and the latter is the EX_2 mechanism (Hvidt & Nielsen, 1966). EX_1 behavior is usually observed at pH values and/or temperatures where ΔG_D is small (Englander & Kallenbach, 1984; Roder et al., 1985).

In many proteins, including cytochromes *c*, most amide protons exchange via the EX_2 mechanism (Gooley et al., 1992; Marmorino et al., 1993; Zhang & Smith, 1993; Bai et al., 1995). For this mechanism,

$$k_{obs} = \frac{k_1 k_{INT}}{k_2} = K_{OP} k_{INT} \quad (3)$$

where K_{OP} is the equilibrium constant for the *N* to *OP* transition. Values of k_{INT} are calculated from sequence specific tripeptide HD exchange data (Molday et al., 1972; Bai et al., 1993; Connelly et al., 1993). We call these calculated values k_{CALC}. The key assumption is that the peptides mimic the open states (i.e., $k_{INT} = k_{CALC}$). The ratio k_{INT}/k_{obs} is *P*, the protection factor. In this model, *P* equals $(K_{OP})^{-1}$, and the local stability or opening free energy, ΔG_{OP}, is $2.3RT \log P$, where *R* is the gas constant and *T* is the absolute temperature.[2]

Investigation of the free energy difference between the native and locally-unfolded states provides insight into the disulfide's effect on open and denatured states. To this end, comparison of ΔG_{OP} to ΔG_D for each protein is used to assign residues to exchange types (i.e., $\Delta G_{OP} < \Delta G_D$, $\Delta G_{OP} \approx \Delta G_D$, or $\Delta G_{OP} > \Delta G_D$). Values of ΔG_{OP} for the wt protein and the cross-linked variant are compared using the relationship

$$\Delta\Delta G_{OP} = \Delta G_{OP,V20C} - \Delta G_{OP,wt} = RT \ln\left(\frac{k_{obs,wt}}{k_{obs,V20C}} \frac{k_{INT,V20C}}{k_{INT,wt}}\right) \quad (4)$$

Further insight is attained by comparing the local stability change to the global stability change using the ratio (Clarke et al., 1993)

$$\frac{\Delta\Delta G_{OP}}{\Delta\Delta G_D} = \frac{\Delta\Delta G_{OP}}{\Delta G_{D,V20C} - \Delta G_{D,wt}} \quad (5)$$

Our current study is aimed at describing the unusual destabilizing effects of a novel disulfide. We have exploited methods that probe the structure and energetics of native and non-native states. Our approach is generally applicable to variants of many different proteins.

MATERIALS AND METHODS

Protein Preparation, pH Measurement, and Global Kinetics. The V20C variant of iso-1-ferricytochrome *c* was prepared as described by Betz and Pielak (1992). The samples were >95% ferricytochrome *c* as determined by visible spectroscopy. DTNB titration (Habeeb, 1972) was used to confirm the presence of the disulfide. pH values are direct readings, uncorrected for any D_2O isotope effect (Glasoe & Long, 1960; Schowen & Schowen, 1982). A manual mixing experiment with Soret circular dichroism detection (Doyle et al., 1996) was used to estimate k_2. The wild-type protein in 1.58 M guanidinium chloride, pH 4.6, where the protein is 98% denatured, was diluted with pH 4.6 buffer to yield a final denaturant concentration of 53 mM, where the protein is >98% native. Kinetics were monitored at 415 nm.

Thermal Denaturation. CD-detected thermal denaturations were performed as described by Cohen and Pielak (1995). To ensure complete exchange before denaturation in D_2O, samples were placed in 50 mM NaOAc-d_3, pH ~7.5, for 25 h at 37 °C. The samples were then lyophilized and stored at −70 °C. The lyophilized protein was resuspended in NaOAc-d_3 to yield a final protein concentration of 20−30 μM and a buffer concentration of 0.06 M, pH 4.6. Values of T_m (the temperature at which half of the protein molecules are denatured) and ΔH_m (the van't Hoff denaturation enthalpy at T_m) were obtained as described by Cohen and Pielak (1994).

NMR Assignments. Data were collected on a Bruker AMX500 spectrometer, transferred to a Silicon Graphics Personal Iris workstation, converted from Bruker to Felix format, and analyzed with Felix 2.0 software. Spectra were acquired using the method of Marion and Wüthrich (1983).

For proton assignment, the concentrations of the ferri and ferro samples were 3 and 7 mM, respectively, in 90% H_2O: 10% D_2O at pH 7.2. NOESY, DQF-COSY, and TOCSY spectra were recorded, collecting 512 increments of 2k points at 300 K, using a sweep width of 8064.52 Hz. Mixing times were 133 ms for the NOESY experiment and 50 ms for the TOCSY experiment. Resonance assignments were made by inspection and comparison to values for the wt protein (Gao et al., 1990) and equine (Feng et al., 1990) cytochrome *c*.

HD Exchange. Samples were prepared as described by Marmorino et al. (1993). The concentration of the ferri and ferro samples were 4 and 3 mM, respectively. Correlated (COSY) spectra were acquired at 298 K and pH 4.6. Spectra of the oxidized protein were recorded at 0.00, 6.55, 13.1, 19.7, 26.2, 32.8, 39.3, 45.9, 52.4, 166, and 243 h. Data were processed as described by Marmorino et al. (1993) except that different internal standards were used. For the ferriform of the cross-linked variant the internal standard is the

[2] Later in this paper we will question the assumption that k_{INT} equals k_{CALC}. Therefore, remember that ΔG_{OP}, K_{OP}, and *P* are *apparent* values.

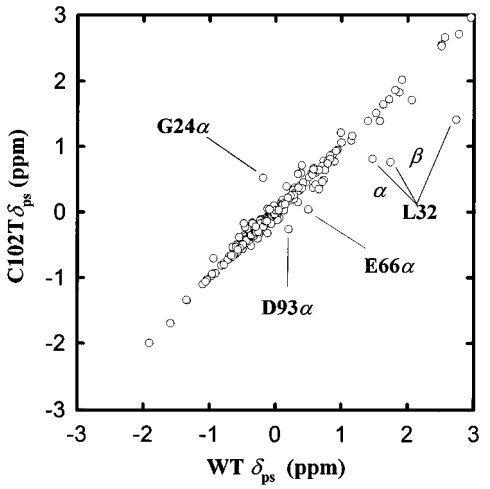

FIGURE 1: Comparison of the paramagnetic shifts from −3 to +3 ppm of all commonly assigned protons of the cross-linked and wild-type proteins.

average of the Leu-15 α-β', Pro-76 α-β, Pro-76 α-β', and Leu-32 α-β' cross-peaks.

RESULTS

Native State Structural Comparisons of Cross-Linked and wt Proteins. To facilitate comparisons with the wt protein (Gao et al., 1990), the cross-linked variant's proton spectrum was extensively assigned at pH 7.2 in both oxidation states. Over 300 assignments are available in each oxidation state of each protein (available as supporting information). We then compared assignments of the cross-linked and wt proteins. Most chemical shift differences between the proteins, are <0.05 ppm (the uncertainty in measuring two chemical shifts is ∼0.02 ppm) and all but a few are <0.2 ppm. The differences are typical of those observed between wt eukaryotic cytochromes *c* (Feng et al., 1990; Gao et al., 1990), supporting the hypothesis that the cross-link does not significantly alter the native state (Betz & Pielak, 1992).

Next we compared δ_{ps} values, a more sensitive probe of structural changes. There are 274 δ_{ps} values common to both proteins. A plot of δ_{ps} between ±3 ppm is presented in Figure 1. To facilitate comparisons we define the parameter Δ_{para} as $\delta_{ps,V20C} - \delta_{ps,wt}$. Protons with $\Delta_{para} \geq |0.40|$ ppm are from Gly-24, Gly-29, Leu-32, Glu-66, Asp-93, and heme propionate proton 17[1a]. The only protons with $\delta_{ps} > |3|$ ppm and a $\Delta_{para} > |0.40|$ ppm are propionate proton 17[1a] (Δ_{para}, −0.59 ppm; $\delta_{ps,wt}$, −8.82 ppm) and a Gly-29 α proton (Δ_{para}, −0.45; $\delta_{ps,wt}$, 3.68 ppm).

We define significant Δ_{para} values as $>|0.08|$ ppm (twice the uncertainty in comparing four chemical shifts) and $>0.3|\delta_{ps,wt}|$. The latter criterion screens out small changes in large $\delta_{ps,wt}$ values. Table 1 lists protons with significant Δ_{para} values. Inspection of the data in Table 1 shows that residues with significant values are distributed throughout the protein.

To ensure that the pH difference between the HD exchange and paramagnetic shift experiments does not affect our conclusions, we compared the chemical shift of amide and α protons of the cross-linked variant at both pH values. All chemical shifts are within 0.05 ppm except the amide (0.10 ppm) and α (0.10 ppm) protons of Trp-59 and the α (0.08 ppm) proton of Asp-93. We also examined Δ_{para} values for the above listed protons at pH 4.6. The pH difference does not account for large Δ_{para} values. In summary, there is no large pH dependent structural change in either protein.

Table 1: Values of $|\Delta_{para}| > 0.08$ ppm and $> |0.3\delta_{ps,wt}|$ (in ppm)

resonance	Δ_{para}	$\delta_{ps,wt}$	resonance	Δ_{para}	$\delta_{ps,wt}$
Phe−3 NH	−0.15	0.09	Met-64 NH	0.11	0.21
Gly-6 α	0.12	0.35	α	−0.13	−0.40
Phe-10 δ	0.09	−0.67	Glu-66 α	−0.46	−0.04
Leu-15 NH	−0.12	−0.77	Tyr-67 α	0.20	−0.59
Gly-24 NH	−0.19	−0.42	Leu-68 α	0.31	0.17
α′	0.71	−0.52	Ile-75 NH	0.22	−1.21
α^2	0.22	−0.39	Gly-77 α′	−0.31	−0.35
Val-28 NH	−0.24	−0.49	Asp-93 α	−0.46	0.26
Gly-29 α′	0.45	6.91	Tyr-97 β	−0.20	0.32
Leu-32 α	−0.66	−0.81	β'	−0.12	0.18
β	−1.32	−1.41	Leu-98 NH	0.17	0.38
β'	−0.98	−0.75	α	0.18	0.17
Gly-37 NH	0.17	−0.05	Glu-103 NH	0.12	−0.10
Glu-60 NH	0.23	0.23	β'	0.13	−0.04
Glu-61 β	0.11	0.11			
β^1	−0.10	−0.03			

Global Stability. For both proteins the ratio of van't Hoff to calorimetric enthalpies is unity, indicating that both proteins denature via a two-state process (Betz & Pielak, 1992; Cohen & Pielak, 1994). To compare global stabilities under HD exchange conditions, CD-detected thermal denaturation experiments were performed in acetate-buffered H_2O and D_2O. Plotting ΔH_m versus T_m for the ferri-forms in H_2O buffer yields a ΔC_p of 1.0 ± 0.2 and 1.4 ± 0.2 kcal mol^{-1} K^{-1} for the cross-linked and the wt protein, respectively. These values are within the uncertainty of those previously reported (Betz & Pielak, 1992; Cohen & Pielak, 1994). The ΔH_m and T_m values for the variant and the wt protein at pH 4.6 are 61.7 kcal mol^{-1}, 332.0 K and 79.6 kcal mol^{-1}, 327.6 K, respectively. The uncertainty in ΔH_m and T_m is ±3.9 kcal mol^{-1} and ±1.1 K (Cohen & Pielak, 1994). Using the integrated form of the Gibbs−Helmholtz equation,

$$\Delta G_D = \Delta H_m\left(1 - \frac{T}{T_m}\right) - \Delta C_p\left[(T_m - T) + T\ln\left(\frac{T}{T_m}\right)\right] \quad (6)$$

these data yield ΔG_D values for the cross-linked and wt protein at pH 4.6 and 298 K of 4.1 ± 0.4 and 5.3 ± 0.4 kcal mol^{-1}, respectively.

To determine stabilities in D_2O, we used T_m values for the fully exchanged proteins in pH 4.6 D_2O buffer and the equation of Becktel and Schellman (1987).

$$\Delta G_{H2O \rightleftharpoons D2O} = \frac{\Delta H_{m,H2O}}{T_{m,H2O}}(T_{m,D2O} - T_{m,H2O}) \quad (7)$$

Values of $\Delta G_{H2O \rightleftharpoons D2O}$ for the cross-linked variant and the wt protein are 1.0 ± 0.4 and 1.1 ± 0.4 kcal mol^{-1}, respectively. The uncertainties in ΔG_D and $\Delta G_{H2O \rightleftharpoons D2O}$ were estimated by applying error propagation analysis to eqs 6 and 7 (Cohen & Pielak, 1994). The ΔG_D increase upon changing solvent was confirmed by examining the pH dependence of ΔH_m and T_m (J. L. Marmorino, unpublished observation). In summary, cross-link introduction decreases stability by 1.2 ± 0.6 kcal mol^{-1} at 298 K, pH 4.6 in H_2O, in agreement with calorimetry data (1.6 kcal mol^{-1}; Betz & Pielak, 1992), and transfer from H_2O to D_2O stabilizes both proteins by 1.0 ± 0.4 kcal mol^{-1}.

Local Stability. We performed HD exchange experiments in deuterated acetate buffer at pH 4.6. Protection factors were calculated as described by Marmorino et al. (1993)

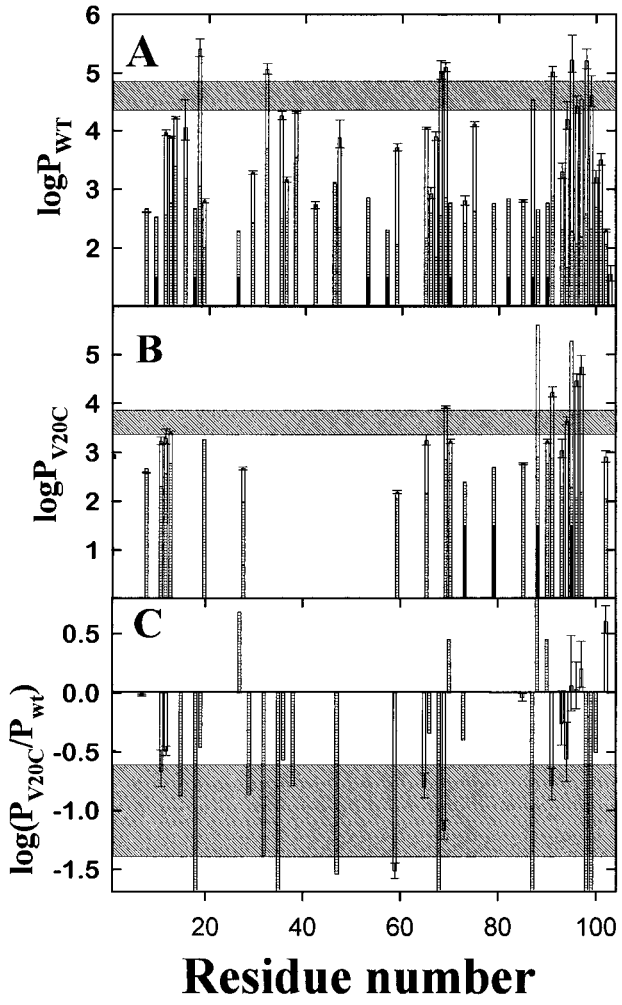

FIGURE 2: Histograms of the logarithm of the protection factor (log *P*) versus residue number, at pH 4.6, 298 K, for the wild-type protein, C102T (A); the cross-linked variant, V20C (B); and $\log(k_{obs,wt}/k_{obs,V20C})$ (C). Inscribed areas are log *P* values corresponding to $\Delta G_{D,D2O} \pm$ one standard deviation for the wild-type protein (A), the cross-linked variant (B), and $\log(k_{obs,wt}/k_{obs,V20C})$ (C). Fast or slow estimated values are indicated by a black bar extending to log $P = 1.5$. Estimated minimum values in panels A and B are indicated by the height of the hatched bars (see text). Errors bars are standard deviations of nonlinear least-square fits of individual decay curves.

using the pH dependent algorithm of Englander et al. (1979; Connelly et al., 1993) and the sequence specific modifications of Bai et al. (1993). Protection factor, k_{obs}, and k_{CALC} values are available as supporting information.

In Figure 2, panels A and B show log *P* histograms for each protein. Panel C is panel B minus A. The log *P* values ($\pm$ one standard deviation) corresponding to $\Delta G_{D,D2O}$ (panels A and B) and $\Delta\Delta G_D$ ($\Delta G_{D,D2O,V20C} - \Delta G_{D,D2O,wt}$, panel C) are also shown. Error bar values are calculated from the standard deviations of k_{obs} from monoexponential fits to HD exchange data. Absence of a bar indicates that either the residue is a proline, the proton is unassigned, the proton signal decays non-monoexponentially, the cross-peak overlaps with another cross-peak, or that the cross-peak disappears before the first time point.

Specific approximations were used where k_{obs} could not be accurately measured. For amide protons whose cross-peaks were present only at the first time point, a k_{obs} of 0.4 h^{-1}, our estimated fast detection limit (Marmorino et al., 1993), is assigned. Three amide protons (Lys-87 and Tyr-97 in the wt protein and Ile-95 in the cross-linked variant) exchanged so slowly that k_{obs} is less than or equal to the standard deviation of the fit. For these residues only a slow detection limit is used. This limit was obtained by adding the standard deviation of the fit to k_{obs}, for each residue. The Glu-88 amide proton exhibited no exchange in the cross-linked variant and its k_{obs} is assumed to be $\leq 0.4 \times 10^{-4}$ h^{-1}, the smallest estimated value. In summary, the fast estimate gives a maximum estimated log *P* and the slow estimate gives a minimum estimated log *P*.

Positions where these fast or slow estimates were made are indicated by a solid bar to log $P = 1.5$. In panels A and B, we assigned estimated log *P* values on the basis of estimated k_{obs} values. These estimates are indicated by the hatched bars. A hatched bar is used in panel C where log *P* for a particular proton is known in one protein but estimated in the other. The size of the bar is the minimum detectable difference. Along the sequence, the change in log *P* between the two proteins varies, but as expected, most log *P* values are smaller in the cross-linked variant. We observe similar patterns for the ferro-forms (data not shown).

DISCUSSION

Native State Structural Differences Are Small. We used changes in paramagnetic shift to assess structural differences. The data support our CD-based conclusions (Betz & Pielak, 1992) that differences between the native states of the variant and the wt proteins are small (Figure 1, Table 1, and Results). For example, His-18 and Met-80 have δ_{ps} values >25 ppm, yet all but one proton on these residues have a Δ_{para} value <0.1 ppm. The exception, His-18 C-5, has a huge δ_{ps} (−23.72 ppm) and a Δ_{para} of only 0.4 ppm. These Δ_{para} values fall through our screen ($\Delta_{para} > |0.08|$ ppm and $> 0.3|\delta_{ps,wt}|$) because δ_{ps} is so large. The small Δ_{para} values indicate little change in chemical environment. The only region that may have a significant structural difference is that near Leu-32. In an iso-1-/iso-2-cytochrome *c* composite protein (Murphy et al., 1992), this region rearranges due to crowding imposed by non-native interactions between positions 20 and 102. Taken together, the data presented here also show that structural changes are not concentrated in any particular region of the protein, although the paramagnetic shift screen ($\Delta_{para} > |0.08|$ ppm and $> 0.3|\delta_{ps,wt}|$) captures many residues (e.g., Leu-32, Met-64, Leu-68, Leu-98) from the leucine rich hydrophobic pocket (Louie et al., 1988; Louie & Brayer, 1990; Lo et al., 1995).

HD Exchange Characteristics. In eukaryotic cytochromes *c*, the amide protons in the N- and C-terminal helices are well-protected against HD exchange (Wand et al., 1986; Paterson et al., 1990; Mayne et al., 1992; Marmorino et al., 1993; Bai et al., 1994, 1995). Examination of all known eukaryotic cytochrome *c* structures shows that the N-terminal helix only interacts with the C-terminal helix (Berghuis & Brayer, 1992). The C-terminal helix, however, interacts with the N-terminal helix, the sixties helix, and the leucine-rich hydrophobic pocket (Lo et al., 1995). In addition to the native state, the amide protons in these helices are also protected in the molten globule (Jeng et al., 1990). Interactions between the N- and C-terminal helices are present in an early folding intermediate of equine cytochrome *c* (Roder et al., 1988; Wu et al., 1993; Sosnick et al., 1994) and stabilize the molten globule (Marmorino & Pielak, 1995). Studies of individual peptides corresponding to each helix show that the isolated C-terminal helix is more stable (i.e.,

helical) than the N-terminal and sixties helices (Kuroda, 1993). Greater C-terminal helix stability correlates with the observation that the amide protons in this helix are more protected than those in the N-terminal helix (Marmorino et al., 1993). This greater stability also correlates with the increased burial of the C-terminal helix in the protein. The isolated N-terminal helix, extended to contain the heme and its His-18 ligand, is more stable than the isolated N-terminal helix, suggesting that the heme directs N-terminal helix formation (Wu et al., 1993).

The engineered disulfide connects the extended N-terminal helix to the end of the C-terminal helix. Qualitative effects of cross-link introduction on HD exchange are seen by comparing log *P* profiles (Figure 2A,B). For the majority of residues, log *P* decreases in the cross-linked variant. Two alternate conclusions can be drawn from this observation. If equivalent amide protons in both proteins exchange from the same open state, it is easier to reach that state in the cross-linked variant. Alternatively, the cross-link may change the open state (i.e., $k_{INT,V20C} \neq k_{INT,wt}$). A more quantitative comparison of Figure 2A,B is given in Figure 2C. Note that the disulfide increases ΔG_{OP} near the C-terminal helix end while decreasing ΔG_D. To understand this and other unusual behavior, we carefully consider ways to classify the results and the assumptions made in the analysis.

Assigning Open States. Analysis of this unusual behavior involves three steps. First, to calculate ΔG_{OP} values, we assume the EX_2 mechanism and that $k_{INT} = k_{CALC}$. Second, we assign open states using ratios of ΔG_{OP} to ΔG_D. Third, in the next section, we examine our conclusions based on both the validity of the assumptions and the possible invalidity of the assumptions.

We divide the open states into three types based on ratios of ΔG_{OP} to ΔG_D. At pH 4.6 and 298 K, $\Delta G_{D,D2O}$ ($\Delta G_{D,H2O} + \Delta G_{H2O \rightleftharpoons D2O}$) is 6.3 ± 0.5 kcal mol^{-1} for the wt protein. The global value, $\Delta G_{D,D2O}$, is converted to log K_D^{-1} (K_D is the denaturation equilibrium constant) for comparison to individual log *P* values. Thus the $\Delta G_{D,D2O}$ value of 6.3 ± 0.5 kcal mol^{-1} corresponds to a log *P* of 4.5 ± 0.4. For the wt protein (Figure 2A), amide protons that exchange by *local* unfolding possess log *P* values between 4.1 and 0. Amide protons that exchange by *global* unfolding possess log *P* values between 4.1 (4.5 − 0.4) and 4.9 (4.5 + 0.4). Amide protons that exchange by *supraglobal* unfolding possess log *P* values >4.9. For the cross-linked protein, $\Delta G_{D,D2O}$ is 5.1 ± 0.5 kcal mol^{-1}, corresponding to a log *P* of 3.6 ± 0.4. Therefore, local, global, and supraglobal unfolding correspond to amide protons with log log *P* values of <3.2, 3.2−4.0, and >4.0. To simplify discussion, we refer to locally unfolded, globally unfolded and super globally unfolded open states as local, global, and supraglobal states.

Our open state nomenclature ranks ΔG_{OP} relative to ΔG_D. The nomenclature does not, however, indicate the position of the open states along the denaturation reaction coordinate. Thus, open states may correspond to locally unfolded states, the denatured state, or states higher in free energy than the denatured state (Marmorino et al., 1993).

Table 2 comprises data from the 36 positions for which data are available for both the cross-linked and wt proteins. Of these 72 protons, 45 exchange from local states, 13 exchange from global states, and 14 exchange from supraglobal states.[3]

Table 2: Local and Global Unfolding Data for Positions with HD Data in Both the Wild-Type and Cross-Linked Proteins

residue	$\Delta\Delta G_{OP}/\Delta\Delta G_D$	open state[a]	residue	$\Delta\Delta G_{OP}/\Delta\Delta G_D$	open state[a]
Ala-7	0.2 ± 0.1	L→L[b]	Leu-68	2.9−2.9	SG→L
Lys-11	0.8 ± 0.4	L→G[b]	Thr-69	1.4 ± 0.7	SG→SG
Thr-12	0.6 ± 0.3	L→G	Asn-70	−0.5 to 3.8	L→G
Leu-15	1.0−3.7	L→L	Lys-73	0.5−2.8	L→L
His-18	2.8−3.6	SG[b]→L	Leu-85	0.0 ± 0.04	L→L
Thr-19	−0.5 to 2.3	L→L	Lys-87	2.5−2.8	SG→L
Thr-19	−0.5 to 2.3	L→L	Glu-88	−6.5 to 3.4	L→SG
Lys-27	−0.8 to 2.3	L→L	Asp-90	−0.5 to 3.8	L→G
Gly-29	1.0−2.8	L→L	Arg-91	0.9 ± 0.5	SG→SG
Leu-32	1.6−4.3	SG→G	Asp-93	0.3 ± 0.4	L→L
Ile-35	2.1−2.9	G→L	Leu-94	0.7 ± 0.5	G→G
Phe-36	0.7−3.0	L→L	Ile-95	−0.1 ± 0.5	SG→SG
Arg-38	0.9−4.1	G→G	Thr-96	0.0 ± 0.3	G→SG
Gln-42	0.0−3.2	L→L	Tyr-97	−0.2 ± 0.3	G→SG
Ser-47	1.8−2.7	L→L	Leu-98	3.0−3.1	SG→L
Trp-59	1.8 ±0.9	L→L	Lys-99	2.7−2.7	G→L
Ser-65	0.9 ± 0.5	L→L	Lys-100	0.6−3.1	L→L
Glu-66	0.4−3.0	L→L	Xxx-102	−0.7 ± 0.1	L→L

[a] Open state in the wild-type protein → open state in cross-linked protein. [b] L, locally unfolded open state; G, globally unfolded open state; SG, superglobally unfolded open state.

Nature of Supraglobal Exchange. Six amide protons in the cross-linked protein (Thr-69, Glu-88, Arg-91, Ile-95, Thr-96, Tyr-97) and eight in the wt protein (His-18, Leu-32, Leu-68, Thr-69, Lys-87, Arg-91, Ile-95, Leu-98) exchange from supraglobal states. This superprotection could arise from the invalidity of the assumption $k_2 \gg k_{INT}$. That is, if exchange occurs via the EX_1 mechanism, then $k_{obs} = k_1$ and it would be inappropriate to interpret *RT* log *P* as ΔG_{OP}. Therefore, we estimated k_2 to assign more confidently the EX_2 mechanism. Folding is complete in <4 s, the manual mixing dead time. We estimate that k_2 is *at least* 1800 h^{-1} and is probably much faster. Thus, k_2 is larger than all k_{CALC} values save one (see supporting information). The situation is probably even more favorable because we compare a minimum k_2 to maximum k_{CALC} values. This observation bolsters our case of the EX_2 mechanism. Although definitive experiments (i.e., HD exchange-directed−quenched flow NMR, mass spectroscopy, and NOE experiments, pH/temperature dependence of exchange) have not been performed, our conclusion is consistent with the fact that equine ferricytochrome *c* amide proton exchange occurs by the EX_2 mechanism (Zhang & Smith, 1993; Bai et al., 1995). In summary, we assume that iso-1-ferricytochrome *c* undergoes HD exchange via the EX_2 mechanism.

Exchange from supraglobal states is unusual because it implies the existence of states higher in free energy than the denatured state (Marmorino et al., 1993). Alternatively, large ΔG_{OP} values may reflect residual open state structure. Such a situation would artificially increase ΔG_{OP} because k_{CALC} values would be $> k_{INT}$ (i.e., the tripeptide model inadequately describes the open states). This idea is also offered to explain the most slowly exchanging protons in the hydrophobic core of lysozyme (Pedersen et al., 1993) and RNase A (Wang et al., 1995). Likewise, the superprotected amide protons in iso-1-ferricytochrome *c* (e.g., His-18, Leu-32, Leu-68, Ile-

[3] Superprotection has also been observed for BPTI (Roder, 1989), RNase A (Mayo & Baldwin, 1993; Wang et al., 1995), and the IgG binding domains of protein G (Orban et al., 1995) but has not been observed for barnase (Perrett et al., 1995), staphylococcal nuclease (Loh et al., 1993), or equine cytochrome *c* (Bai et al., 1995).

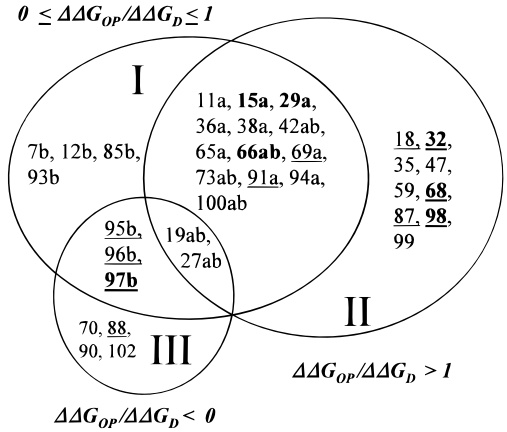

FIGURE 3: Venn diagram summarizing data for class I, II, and III residues. Arabic numbers refer to sequence position and lower case letters subdivide class I. For class Ia $\Delta\Delta G_{OP}/\Delta\Delta G_D$ is within uncertainty of unity, and for class Ib $\Delta\Delta G_{OP}/\Delta\Delta G_D$ is less than 1. Due to experimental uncertainties some class I residues fall into both subdivisions. Bolded entries indicate that a proton or protons from that residue are captured by the paramagnetic shift screen (Table 1), and underlined entries indicate that these amide protons exhibit superprotection in one or both proteins.

95, Leu-98) comprise the leucine-rich hydrophobic pocket (Louie et al., 1988; Lo et al., 1995).

There is also evidence of slow exchange in denatured cytochromes *c* from studies of chemically denatured states. Under conditions where the His-18 imidazole of equine cytochrome *c* remains ligated to the iron, log *P* for its amide proton is 2.1 (Muthukrishnan & Nall, 1991; Elöve et al., 1994). Thus, there is residual structure near His-18 in the open state of urea-denatured equine ferricytochrome *c*. His-18 in iso-1-ferricytochrome *c* is also superprotected (Table II, Figure 2). Furthermore, the amide protons from Arg-13, His-18, Leu-32, Phe-36, Thr-69, and Arg-91 of iso-1-ferricytochrome *c* are protected in the chemically denatured state (Betz, 1993a). These data show the existence of structure in non-native states.

In summary, the observation $\Delta G_{OP} > \Delta G_D$ may reflect residual open state structure that is absent in the peptides from which k_{CALC} is obtained (i.e., $k_{INT} \neq k_{CALC}$). This potential problem of open state structure is consistent with the observation that folding intermediates often exist under native state conditions. To overcome and gain information about this potential problem, we analyzed the data in terms of the ratio $\Delta\Delta G_{OP}/\Delta\Delta G_D$ (eq 5). In this case the tripeptide data are not necessary if we assume $k_{CALC,V20C}/k_{CALC,wt} = k_{INT,V20C}/k_{INT,wt}$ (eq 4). In fact, for all but four residues (those bracketing the substitutions), this assumption reduces to assuming that the ratios are unity [i.e., $\Delta\Delta G_{OP} = RT \ln(k_{obs,wt}/k_{obs,V20C})$].

Classifying Open States. Energetic effects of cross-link introduction are examined by comparing $\Delta\Delta G_{OP}/\Delta\Delta G_D$ (eq 5) to the identity (local, global, or supraglobal)—or change of identity—of the open state (Table 2). When ΔG_{OP} is defined in one protein but not the other, a $\Delta\Delta G_{OP}/\Delta\Delta G_D$ range was calculated. One extreme is calculated from the estimated log *P* (see Results). The other is calculated by assuming that the missing log *P* is zero.

To analyze the data we define three open state classes. A Venn diagram showing the classifications is presented in Figure 3. Position numbers in bold are captured by the paramagnetic shift screen ($\Delta_{para} > |0.08|$ ppm and $> 0.3|\delta_{ps,wt}|$, Table 1). Underlined position numbers are superprotected (Table 2). Some residues fall into more than one class because of experimental uncertainty and the fact that we used a range of estimates for some residues.

Class I is defined by the criterion $1 \geq \Delta\Delta G_{OP}/\Delta\Delta G_D \geq 0$ and is subdivided as described by Clarke et al. (1993). In class Ia, $\Delta\Delta G_{OP}/\Delta\Delta G_D$ is within uncertainty of unity, indicating that the perturbation affects ΔG_D and ΔG_{OP} equally. That is, these residues realize the full effect of the perturbation. In class Ib, $\Delta\Delta G_{OP}/\Delta\Delta G_D$ is <1 but >0. For these residues, open state free energy changes are only weakly coupled to $\Delta\Delta G_D$. In class II, $\Delta\Delta G_{OP}/\Delta\Delta G_D$ is greater than unity. That is, the opening free energy decrease is greater than the stability decrease, an unusual observation. In class III, $\Delta\Delta G_{OP}/\Delta\Delta G_D$ is negative. That is, the cross-link increases local stability while decreasing global stability, another unusual observation.

Residues exclusive to class I are not captured by the paramagnetic shift screen. Therefore, the native state structural and energetic effects of the cross-link are negligible for class I residues.

As expected, class Ia ($\Delta\Delta G_{OP}/\Delta\Delta G_D \approx 1$) and class II ($\Delta\Delta G_{OP}/\Delta\Delta G_D > 1$) show the greatest overlap. With one exception (Tyr-97), residues captured by the paramagnetic shift screen belong to class II. That is, $\Delta\Delta G_{OP}$ is related to cross-link induced native state structural change(s) for class II residues. There are four additional important features. First, all exclusively class II residues captured by the paramagnetic shift screen (Leu-32, Leu-68, Leu-98) exchange from a supraglobal state. Second, these residues form the leucine-rich hydrophobic core (Lo et al., 1995). Third, comparison of Table 2 and Figure 3 shows that for these three residues, cross-link introduction causes the open state to decrease in free energy relative to the native state (i.e., demotion from supraglobal to global or local states). Fourth, class II residues His-18, Leu-32, Phe-36, Thr-69, and Arg-91 are protected from exchange in the chemically denatured state of the wt protein (Betz, 1993a). Taken together, these observations suggest that for class II residues, $k_{INT,V20C}$ is greater than $k_{INT,wt}$. That is, these open states change on cross-link introduction (eq 4). Furthermore, the increased $k_{INT,V20C}$ values are linked to hydrophobic core structural changes. We conclude that for class II residues the cross-link relaxes native-like hydrophobic core structure that is present in high-energy open states of the wt protein. A similar conclusion is reported for mutation-induced changes in protein G domains (Orban et al., 1995).

Class III ($\Delta\Delta G_{OP}/\Delta\Delta G_D < 0$) has the smallest number of residues, but many are close to the cross-link in primary and tertiary structure. For class III residues ΔG_{OP} increases despite the ΔG_D decrease. Such increased protection also occurs for amide protons close in sequence to the disulfide of denatured lysozyme (Radford et al., 1992). Three consecutive class III residues (95−97) exchange from a global or *supraglobal* state in both proteins. This means that some of the most well-protected protons are little affected by the cross-link, despite the fact that the cross-link significantly decreases ΔG_D. We conclude that for class III residues, $k_{INT,V20C}$ is less than $k_{INT,wt}$, because the disulfide induces a more structured open state. Evidently the cross-link-induced open state structure is that detected in CD studies of the denatured cross-linked variant (Betz & Pielak, 1992). Furthermore, native state structural changes are not associated with class III residues, because only one (Tyr-97) is captured by the structural screen. Our analysis

suggests that for class III residues, the cross-link induces open state structure without affecting native state structure, a conclusion that complements our earlier work (Betz & Pielak, 1992).

CONCLUSIONS

Introduction of a cross-link into yeast iso-1-ferricytochrome *c* does not dramatically change its native structure.

The majority of amide protons react to introduction of the disulfide as expected in that changes in local stability are less than or equal to the change in global stability.

Some residual structure in high-energy open states of the wild-type protein is rearranged on cross-link introduction. This rearrangement involves relaxation of open state hydrophobic core structure and induction of structure near the disulfide.

Induced open state structure near the disulfide may correspond to residual helical structure found in the CD spectrum of the cross-linked variant's denatured state (Betz & Pielak, 1992).

ACKNOWLEDGMENT

We thank Charles Carter, Jr., Marshall Edgell, Jan Hermans, Jr., Terry Oas, Richard Talbert, and members of the Pielak lab for helpful discussions.

SUPPORTING INFORMATION AVAILABLE

A table of proton NMR assignments for the Fe(II) and Fe(III) forms of the V20C variant and a table of k_{CALC}, k_{obs}, and log *P* values for the V20C variant are available (27 pages). Ordering information is given on any current masthead page.

REFERENCES

Auld, D. S., Young, G. B., Saunders, A. J., Doyle, D. F., & Pielak, G. J. (1993) *Prot. Sci. 2*, 2187−2197.
Bai, Y., Milne, J. S., Mayne, L., & Englander, S. W. (1993) *Proteins 17*, 75−86.
Bai, Y., Milne, J. S., Mayne, L., & Englander, S. W. (1994) *Proteins 20*, 4−14.
Bai, Y., Sosnick, T. R., Mayne, L., & Englander, S. W. (1995) *Science 269*, 192−197.
Baldwin R. L. (1993) *Curr. Opin. Struct. Biol. 3*, 84.
Becktel, W. J., & Schellman, J. A. (1987) *Biopolymers 26*, 1859−1877.
Berger, A., & Linderstrøm-Lang, K. U. (1957) *Arch. Biochem. Biophys. 69*, 106−118.
Berghuis, A. M., & Brayer, G. D. (1992) *J. Mol. Biol. 223*, 959−976.
Betz, S. F. (1993a) Doctoral Dissertation, University of North Carolina, Chapel Hill.
Betz, S. F. (1993b) *Protein Sci. 2*, 1511−1518.
Betz, S. F., & Pielak, G. J. (1992) *Biochemistry 31*, 12337−12344.
Brems, D. N., Cass, R., & Stellwagen, E. (1982) *Biochemistry 21*, 1488−1493.
Bryant, C., Strottmann, J. M., & Stellwagen, E. (1985) *Biochemistry 24*, 3459−3464.
Clarke, J., Hounslow, A. M., Bycroft, M., & Fersht, A. R. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 9837−9841.
Cohen, D. S., & Pielak, G. J. (1994) *Protein Sci. 3*, 1253−1260.
Cohen, D. S., & Pielak, G. J. (1995) *J. Am. Chem. Soc. 117*, 1675−1677.
Connelly, G. P., Bai, Y., Jeng, M. F., & Englander, S. W. (1993) *Proteins: Struct., Funct., Genet. 17*, 87−92.
Cutler, R. L., Pielak, G. J., Mauk, A. G., & Smith, M. (1987) *Protein Eng. 1*, 95−99.
Doyle, D. F., Waldner, J. C., Parikh, S., Alcazar-Roman, L., & Pielak, G. J. (1996) *Biochemistry 35* (in press).
Elöve, G. A., Bhuyan, A. K., & Roder, H. (1994) *Biochemistry 33*, 6925−6935.
Englander, S. W., & Kallenbach, N. R. (1984) *Q. Rev. Biophys. 16*, 521−655.
Englander, J. J., Calhoun, D. B., & Englander, S. W. (1979) *Anal. Biochem. 92*, 517−524.
Feng, Y., Roder, H., & Englander, S. W. (1990) *Biochemistry 29*, 3494−3504.
Flory, P. J. (1956) *J. Am. Chem. Soc. 78*, 5222−5235.
Gao, Y., Boyd, J., Williams, R. J. P., & Pielak, G. J. (1990) *Biochemistry 29*, 6944−7003.
Gao, Y., Boyd, J., Pielak, G. J., & Williams, R. J. P. (1991) *Biochemistry 30*, 1928−1934.
Glasoe, P. K., & Long, F. A. (1960) *J. Phys. Chem. 64*, 188−190.
Gochin, M., & Roder, H. (1995) *Protein Sci. 4*, 296−305.
Gooley, P. R., Caffrey, M. S., Cusanovich, M. A., & MacKenzie, N. E. (1992) *Biochemistry 31*, 443−450.
Habeeb, A. F. S. A. (1972) *Methods Enzymol. 25*, 457−464.
Harrison, P. M., & Sternberg, M. J. E. (1994) *J. Mol. Biol. 244*, 448−463.
Hvidt, A., & Nielsen, S. O. (1966) *Adv. Protein Chem. 21*, 287−386.
Jeng, M. F., Englander, S. W., Elöve, G. A., Wand, A. J., & Roder, H. (1990) *Biochemistry 29*, 10433−10437.
Kuroda, Y. (1993) *Biochemistry 32*, 1219−1224.
Lo, T. P., Murphy, M. E. P., Guillemette, J. G., Smith, M., & Brayer, G. D. (1995) *Protein Sci. 4*, 198−208.
Loh, S. N., Prehoda, K. E., Wang, J., & Markley, J. L. (1993) *Biochemistry 32*, 11022−11028.
Louie, G. V., & Brayer, G. D. (1990) *J. Mol. Biol. 214*, 527−555.
Louie, G. V., Hutcheon, W. L. B., & Brayer, G. D. (1988) *J. Mol. Biol. 199*, 295−314.
Marion, D., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 113*, 967−974.
Marmorino, J. L., & Pielak, G. J. (1995) *Biochemistry 34*, 3140−3143.
Marmorino, J. L., Auld, D. S., Betz, S. F., Doyle, D. F., Young, G. B., & Pielak, G. J. (1993) *Protein Sci. 2*, 1966−1974.
Mathews, F. S. (1985) *Prog. Biophys. Mol. Biol. 45*, 1−56.
Mayne, L., Paterson, Y., Cerasoli, D., & Englander S. W. (1992) *Biochemistry 31*, 10678−10685.
Mayo, S. L., & Baldwin, R. L. (1993) *Science 262*, 873−876.
Moore, G. R., & Pettigrew, G. W. (1990) *Cytochromes c: Evolutionary, Structural and Physiochemical Aspects*, Springer-Verlag, Berlin.
Molday, R. S., Englander, S. W., & Kallen, R. G. (1972) *Biochemistry 11*, 150−158.
Murphy, M. E. P., Nall, B. T., & Brayer, G. D. (1992) *J. Mol. Biol. 227*, 160−176.
Muthukrishnan, K., & Nall, B. T. (1991) *Biochemistry 30*, 4706−4710.
Orban, J., Alexander, P., Bryan, P., & Khare, D. (1995) *Biochemistry 34*, 15291−15300.
Paterson, Y., Englander S. W., & Roder, H. (1990) *Science 249*, 755−759.
Pedersen, T. G., Thomsen, N. K., Andersen, K. V., Madsen, J. C., & Poulsen, F. M. (1993) *J. Mol. Biol. 230*, 651−660.
Perrett, S., Clarke, J., Hounslow, A. M., & Fersht, A. R. (1995) *Biochemistry 34*, 9288−9298.
Radford, S. E., Buck, M., Topping, K. D., Dobson, C. M., & Evans, P. A. (1992) *Proteins 14*, 237−248.
Roder, H. (1989) *Methods Enzymol. 176*, 446−473.
Roder, H., Wagner, G., & Wüthrich, K. (1985) *Biochemistry 24*, 7396−7407.
Roder, H., Elöve, G. A., & Englander, S. W. (1988) *Nature (London) 335*, 700−704.
Schowen, K. B., & Schowen, R. L. (1982) *Methods Enzymol. 87*, 551−606.
Sosnick, T. R., Mayne, L., Hiller, R., & Englander, S. W. (1994) *Nat. Struct. Biol. 1*, 149−156.
Wand, A. J., Roder, H., & Englander, S. W. (1986) *Biochemistry 25*, 1107−1114.
Wang, A., Robertson, A. D., & Bolen, D. W. (1995) *Biochemistry 34*, 15096−15104.
Wu, L. C., Laub, P. B., Elöve, G. A., Carey, J., & Roder, H. (1993) *Biochemistry 32*, 10271−10276.
Zhang, Z., & Smith, D. L. (1993) *Protein Sci. 2*, 522−531.

BI9528558