The Influence of Intramolecular Bridges on the Dynamics of a Protein Folding Reaction[†]

Jody M. Mason,*,‡,§ Nicholas Gibbs,‡ Richard B. Sessions,‡ and Anthony R. Clarke‡,ǁ

Department of Biochemistry, School of Medical Sciences, University of Bristol, Bristol, BS8 1TD, U.K., and MRC Prion Unit, Department of Neurogenetics, Imperial College School of Medicine at St. Mary's, London, W2 1NY, U.K.
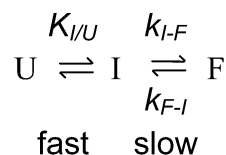
ABSTRACT: Thirteen versions of a β-sheet protein have been constructed, each with a single, surface-exposed disulfide bridge. A comparison of folding kinetics, in oxidizing and reducing conditions, is used to elucidate the order in which β-strands become associated during the folding process and, hence, the relationship between topology and folding dynamics. In common with the wild-type molecule, all the proteins fold through a two-step (three state) mechanism with a rapidly formed intermediate which slowly converts to the native state. In a majority of cases, the bridge is seen to stabilize the folded state, and for five of the modified proteins, the additional stability is greater than 3 kcal/mol. Surprisingly, cross-links which connect β-strands which are distant in sequence predominantly stabilize the rapidly formed intermediate state, suggesting that these strand−strand interactions occur in the initial stages of folding. Cross-links which stabilize local hairpins have their major influence on the second, rate-determining step leading to significant enhancements in the folding rate. We find that enhancement of the folding rate in the second, rate-limiting step is correlated with a reduction in contact order in the same way as in naturally occurring proteins of different folds. The large increases in native-state stability resulting from the insertion of disulfide bridges on the surface of β-sheet structures have implications for enhancing the robustness of proteins by molecular engineering.

A major goal of protein science is to elucidate the sequence of events which occurs when an unstructured polypeptide chain folds to form a compact and unique native conformation. This knowledge is central to our understanding of the way in which sequence determines the fold of protein molecules and, in consequence, has implications for the development of methods for translating genomic data into protein structure and function. Two techniques have contributed most to our experimental understanding of the folding process. First, the influence of point mutations on the energies of defined states in the folding reaction provides information on the degree of nativelike structure acquired at a given locus in the molecule (*1−3*). Second, the kinetics of hydrogen−deuterium exchange of specific amide groups along the backbone of the polypeptide can be used to determine those which participate in hydrogen bonds in populated intermediates (*4*).

In principle, the kinetic influence of well-chosen cross-links in a protein molecule can be used as a further tool to probe the development of chain topology during folding (*5−8*). The differential effects of cross-bridge positions on protein folding have recently been examined theoretically by Abkevich and Shaknovich (*7*), using a lattice model of the folding process. It is timely to compare these findings with those derived from in vitro measurements. For the experimental study described here, we chose the all-β-sheet protein CD2 as a test-bed for disulfide engineering. Shown in Figure 1 is the backbone structure of this immunoglobulin-like domain. Unusually for a protein of this class, it contains no covalent cross-bridges, and so is ideal as a template for this study. Further, the folding kinetics of CD2 are rigorously established (*9−12*), and the principles governing the positioning of disulfide cross-bridges in β-sheet proteins are well-understood (*13*, *14*).

The molecule folds through a rapidly formed, partially folded intermediate which then slowly converts to the native state as described by the following scheme:

$$U \underset{}{\overset{K_{I/U}}{\rightleftharpoons}} I \underset{k_{F\text{-}I}}{\overset{k_{I\text{-}F}}{\rightleftharpoons}} F$$

$$\text{fast} \quad \text{slow}$$

Hence, there are three progressively structured states which can be characterized experimentally; the intermediate state, the major transition state, and the native state. The basis of our analysis is as follows; suppose that the CC′ hairpin (see Figure 1) were formed in the first step of the folding reaction, then bridging this hairpin should promote this step and stabilize the intermediate with respect to the unfolded state. If it is not formed until the major transition state, then it will have no effect on the intermediate but will accelerate the formation of the native state. If it is only formed in the

* Correspondence should be addressed to this author. Tel: +44 161 200 4191. Fax: +44 161 236 0409. Email: J.Mason-3@umist.ac.uk.
‡ University of Bristol.
§ Current address: Department of Biomolecular Sciences, University of Manchester Institute of Science and Technology, Manchester M60 1QD, U.K.
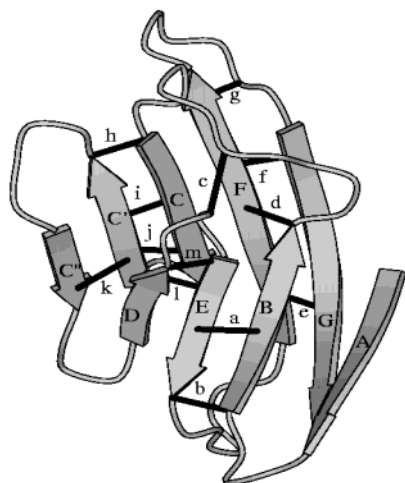ǁ Imperial College School of Medicine at St. Mary's.

FIGURE 1: Positions of disulfide bridges. The topology of the protein is that of an immunoglobulin variable (V) domain with two sheets, the larger containing strands C″C′CFGA, and the smaller DEB. The molecule has nine strands in total with four well-defined hairpins (F−G, D−E, C−C′, and C′-C″), and three sequence-distant strand−strand interactions (B−F, C−F, and B−E). Four sequence-distant cross-links (13−66, 15−64, 23−60, 18−78) and nine sequence-local (hairpin) cross-links (29−43, 31−41, 33−38, 40−50, 58−62, 57−63, 81−87, 79−90, 75−94) were introduced. The bridges are given in lower case letters (a−m) and refer to the residue positions shown in Table 1.

folded state, then the reaction in the direction of folding will be unaffected, but the rate of unfolding will be reduced.

Apart from the above aspects of folding dynamics and topology, there are two further motives for performing this study. First, there has never emerged a thorough understanding of the effect of disulfide bridges on protein stability, as there is as strong dependence on the protein used and the bridge position (*15*, *16*). By surveying a large and representative range of sites in a protein, it is possible that empirical rules might emerge, at least in the case of β-sheet structures, for stabilizing the folded state using disulfide bridges. An understanding of such rules will aid efforts to produce more robust proteins for medical and industrial use by genetic engineering (*5*). Second, there has been much recent interest in the effects of contact order on folding rates for different folded-state topologies. A correlation has been observed between the average sequence-spacing of inter-residue contacts in a protein and the rate of folding. These comparisons have been performed on proteins of different

folds and on proteins of the same fold with altered loop lengths (*17*, *18*). An examination of the effects of bridges provides a further test of these ideas by reducing the contact order to a greater or lesser extent depending upon bridge position. In making these changes in contact order, there is a minimal alteration of sequence so that the relationship between contact order and folding rate in these variants should be evident.

A series of 13 versions of CD2 were constructed, each with a pair of cysteine residues at locations in the sequence where a disulfide bond will readily form in the native state. These locations are typically pairs of surface residues in non-hydrogen-bonding sites of antiparallel β-strands (*13*, *14*). The positions of these bridges in the context of the CD2 structure are shown in Figure 1. The proteins were then subjected to a full kinetic analysis of folding and unfolding (see Figure 2) in order to determine the folding parameters k_{I-F}, k_{F-I}, and $K_{I/U}$ in native conditions. This procedure was performed with the proteins in the reduced state (no bridge) and in the oxidized state (intact bridge) to assess the effect of the bridge on the folding dynamics.

MATERIALS AND METHODS

Mutagenesis. All mutations were made on the rat gene cloned into the pGEX-2T glutathione *S*-transferase (GST)-fusion vector (*29*). The proteins were expressed and purified as described previously (*30*).

Cross-Linking Disulfides. To form the disulfide bridges, 10 μM protein was oxidized in an oxygen-purged buffer containing 20 μM zinc chloride, 50 mM TEA, at pH 7.5. To break the bridges, 10 μM protein in a helium-purged buffer containing 1 mM EDTA and 50 mM TEA at pH 7.5 was reduced by the addition of 10 mM DTT. In addition, all other buffers were filtered and purged with the relevant gases. The proteins were then checked for oxidation and reduction using mass spectrometry and Elman's reagent (*31*).

Kinetic Analysis. The data have been fitted to the equation: $k_{obs} = k_{F-I} + k_{I-F}/(1 + 1/K_{I/U})$ where k_{F-I} and k_{I-F} are rate constants describing the forward and reverse reactions, respectively, between the folded and intermediate states (I). $K_{I/U}$ is the equilibrium constant ([I]/[U]) for the rapid interconversion of the intermediate and unfolded states (*32*). In the fitting routine, the following relationships were used: $k_{F-I} = k_{F-I(w)} \exp(-m_t D)$; $k_{I-F} = k_{I-F(w)} \exp[(m_I - m_t)D]$;
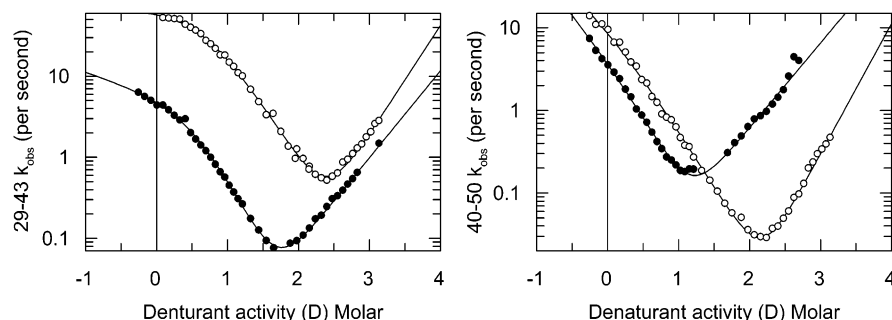

FIGURE 2: Kinetic analysis of the effect of cross-links. Examples of the folding dynamics of two versions of CD2 (29−41, C−C′ and 40−50, C′-C″) are shown. Folding and unfolding reactions were simple first-order processes for all proteins, and the observed relaxation rate constants (k_{obs}) are plotted as a function of denaturant activity as described previously (*30*). The open circles represent the oxidized version (cross-link intact) while the closed circle chevron plots represent the reduced version (cross-link broken). The data illustrate the contrasting effects of these bridges, with that in the CC′ hairpin mainly affecting the folding 'limb' of the plot (i.e., at low denaturant) while that in C′C″ mainly affects unfolding.

Table 1: Equilibrium and Rate Constants for the Disulfide Mutants[a]

		$k_{I-F(w)}$ (s^{-1})	$K_{F-I(w)}$ (s^{-1})	$K_{I/U(w)}$	m_u (M^{-1})	m_I (M^{-1})	m_t (M^{-1})
(a) G13C-K66C EB	ox	4.7 ± 0.5	2.0e-6 ± 1.2e-6	86.7 ± 16.7	−8.6 ± 0.2	−4.0 ± 0.3	−5.3 ± 0.2
	red	1.3 ± 0.2	1.6e-2 ± 6.5e-5	18.0 ± 4.2	−7.3 ± 0.2	−2.1 ± 0.6	−2.8 ± 0.1
(b) N15C-K64C EB	ox	4.2 ± 0.2	1.3e-2 ± 1.3e-2	6.7 ± 1.4	−7.6 ± 0.1	−2.8 ± 0.2	−2.7 ± 0.1
	red	4.5 ± 4.0	5.4e-4 ± 1.2e-4	3.3 ± 0.6	−6.4 ± 0.1	−1.8 ± 0.7	−2.7 ± 0.1
(c) M23C-N60C DE/BC	ox	2.9 ± 0.2	7.5e-5 ± 1.4e-5	5.5 ± 2.1	−6.8 ± 0.2	−4.0 ± 0.2	−3.3 ± 0.1
	red	9.2 ± 0.4	1.0e-2 ± 3.7e-4	0.8 ± 0.1	−5.6 ± 0.0	−2.3 ± 0.2	−2.2 ± 0.0
(d) I18C-I78C BF	ox	2.9 ± 0.7	1.4e-3 ± 2.9e-5	11.5 ± 0.5	−7.3 ± 0.0	−3.5 ± 0.0	−3.1 ± 0.0
	red	2.9 ± 0.0	1.6e-2 ± 6.5e-5	0.6 ± 4.3	−6.0 ± 0.0	−2.3 ± 0.0	−2.0 ± 0.0
(e) T75C-D94C FG	ox	15.4 ± 0.7	8.8e-4 ± 2.0e-4	151.8 ± 65.5	−8.2 ± 0.2	−3.8 ± 0.1	−3.3 ± 0.1
	red	4.2 ± 0.2	1.1e-3 ± 9.4e-5	29.7 ± 7.0	−8.3 ± 0.2	−4.4 ± 0.1	−2.7 ± 0.0
(f) T79C-D90C FG	ox	26.4 ± 0.5	5.8e-3 ± 5.0e-4	83.4 ± 6.8	−7.4 ± 0.0	−2.9 ± 0.0	−2.5 ± 0.0
	red	1.6 ± 0.0	3.1e-3 ± 1.1e-4	12.5 ± 1.0	−7.4 ± 0.0	−2.6 ± 0.0	−2.3 ± 0.0
(g) Y81C-R87C G	ox	2.3 ± 0.1	2.7e-2 ± 1.8e-3	8.2 ± 2.9	−7.2 ± 0.3	−2.9 ± 0.2	−2.0 ± 0.0
	red	1.0 ± 0.1	1.6e-3 ± 1.6e-4	5.3 ± 2.4	−7.6 ± 0.4	−3.6 ± 0.3	−2.4 ± 0.0
(h) E29C-K43C CC′	ox	65.2 ± 3.8	1.7e-4 ± 7.0e-5	7.0 ± 1.5	−6.2 ± 0.1	−3.1 ± 0.3	−3.1 ± 0.1
	red	5.6 ± 0.3	6.0e-4 ± 7.0e-5	4.9 ± 1.3	−6.3 ± 0.1	−3.1 ± 0.2	−2.5 ± 0.0
(i) R31C-E41C CC′	ox	90 ± 0.0	1.7e-3 ± 4.1e-4	5.0 ± 0.5	−6.4 ± 0.1	−2.7 ± 0.3	−2.8 ± 0.1
	red	4.6 ± 0.2	1.8e-3 ± 2.7e-5	10.0 ± 3.0	−7.5 ± 0.2	−4.6 ± 0.2	−3.1 ± 0.1
(j) E33C-I38C CC′	ox	19.6 ± 2.3	3.4e-2 ± 4.5e-3	2.2 ± 0.7	−6.6 ± 0.2	−2.5 ± 0.2	−2.3 ± 0.1
	red	5.7 ± 0.4	1.1e-3 ± 1.6e-4	4.8 ± 1.4	−7.9 ± 0.2	−3.8 ± 0.2	−3.0 ± 0.1
(k) A40C-I50C C′C″	ox	19.5 ± 1.8	5.8e-6 ± 1.5e-7	0.8 ± 0.0	−6.8 ± 0.0	−5.0 ± 0.0	−3.6 ± 0.0
	red	4.9 ± 0.7	5.4e-3 ± 8.8e-4	2.7 ± 1.5	−6.5 ± 0.3	−4.8 ± 0.3	−2.4 ± 0.1
(l) I58C-D62C DE	ox	2.8 ± 0.1	4.1e-3 ± 4.1e-4	34.6 ± 17.5	−10.2 ± 0.5	−4.8 ± 0.1	−2.6 ± 0.1
	red	6.7 ± 0.8	8.3e-5 ± 3.3e-6	3.8 ± 1.9	−9.1 ± 0.3	−5.4 ± 0.4	−4.5 ± 0.1
(m) I57C-I63C DE	ox	54.8 ± 4.5	1.7e-5 ± 3.5e-6	47.8 ± 5.7	−8.2 ± 0.1	−4.3 ± 0.2	−4.1 ± 0.1
	red	4.4 ± 0.2	7.3e-4 ± 9.0e-5	12.7 ± 4.0	−7.9 ± 0.2	−3.3 ± 0.2	−2.9 ± 0.0
wild type	−		6.0 ± 0.3	5.4e-4 ± 1e-4	3.6 ± 0.6	−7.1 ± 0.1	−2.9 ± 0.2	−2.7 ± 0.1

[a] The bridge position is described by the mutation in the single-letter code and by the strands that are joined (by upper case letter as shown in Figure 1). The values for parameters derived from the optimal fits to the three-state rate equation (see Materials and Methods and Figure 2) are shown for data collected in oxidizing (ox) and reducing (red) conditions.

$K_{I/U} = K_{I/U(W)} \exp[(m_u − m_I)D]$, where the subscript w describes the rate and equilibrium constants in water, and the *m* parameters describe the shifts in the stabilities of each state (designated by the subscript) as a function of the denaturant activity (*D*), where $D = [C_{0.5}[GuHCl]/(C_{0.5} + [GuHCl])] − 2.6[Na_2SO_4]$, where $C_{0.5}$ is a denaturation constant with a value of 7.5 M. This treatment has been explained in detail elsewhere (*12, 30, 32*), and the coefficient of 2.6 is derived from the linear relationship between Na_2SO_4 concentration and denaturant activity between 0 and 0.4 M (i.e., solutions of 0, 0.1, 0.2, 0.3, and 0.4 M give molar denaturant activities of 0, −0.27, −0.51, −0.78, and −1.03 M, respectively). Sodium sulfate is a cosmotropic agent and works in a manner analagous but opposite to GuHCl, increasing the free energy of solvation of hydrocarbon and consequently driving the folding reaction in favor of more compact and desolvated states. The molar ability of Na_2SO_4 to decrease the extent of hydrocarbon burial has been scaled to the molar ability of GuHCl to increase it, allowing us to calculate the denaturant activity of Na_2SO_4 (*32*). This results in negative values of denaturant activity.

Folding measurements were initiated by mixing a 10 μM solution of unfolded CD2.d1[1] containing 50 mM TEA and 3.27 M GuHCl against 10 volumes of the given concentrations of GuHCl at 298 K in an SX.18MV stopped-flow apparatus (Applied Photophysics Ltd.). An excitation wavelength of 295 nm was selected by a single monochromator (band-pass 5 nm) from a mercury−xenon light source. The fluorescence intensity above 340 nm was recorded using an emission cutoff filter. For the unfolding reactions, a 10 μM

[1] Abbreviations: CD2.d1, domain 1 of cluster determinant 2; GST, glutathione *S*-transferase.

solution of folded CD2.d1 in 50 mM TEA, pH 7.5, was mixed against 10 volumes of an appropriate concentration of GuHCl at 298 K, and the reaction was recorded as described above. The resulting data points are the result of at least three averages.

Structure. The coordinates for wild-type CD2.d1 were obtained from the Brookhaven Protein DataBank (1HNG).

RESULTS AND DISCUSSION

Effects on Overall Stability. The rate and equilibrium constants for individual steps in the folding/unfolding reaction for the 13 proteins were determined by measuring relaxation rates as a function of denaturant activity as shown for 2 example proteins in Figure 2. These measurements were performed with the proteins in both the reduced and oxidized states. All experimental data are summarized in Table 1. These effects have also been translated to free energy changes associated with each stage in the reaction (Table 2). To assess the energetic influence of forming the bridge, free energy changes are given with reference to the reduced state of that protein rather than with reference to the wild-type. By doing this, the local structural perturbation caused by replacing the wild-type residue with a cysteine is discounted, and we reveal only the effect of oxidizing the cysteines, which will have a minimal influence on hydration and bulk.

Given the variability of success in previous attempts to stabilize proteins by disulfide engineering, it is first worth examining the effect on the overall stability of the molecule, i.e., on ΔG for the U-to-F transition. The stabilizing effect of introducing a bridge into a protein chain is rationalized by the argument that the bridge limits the number of conformations accessible to the unfolded state, thereby

Table 2: Effects of Bridges on the Free Energy of Defined Steps in the Folding Process[a]

bridge position	$\Delta\Delta G_{(U\to F)}$ (predicted)	$\Delta\Delta G_{(U\to F)}$ (measured)	$\Delta\Delta G_{(U\to I)}$	$\Delta\Delta G_{(I\to t)}$	$\Delta\Delta G_{(t\to F)}$	$\Delta G_{(U\to F)}$ (oxidized)
(a) g13c-k66c EB	−4.1	−7 ± 0.4	−0.9 ± 0.1	−0.8 ± 0.1	−5.3 ± 0.3	−11.3 ± 0.36
(b) n15c-k64c EB	−4.1	1.5 ± 0.3	−0.4 ± 0.2	0.0 ± 0.1	1.9 ± 0.2	−4.55 ± 0.6
(c) m23c-n60c DE/BC	−3.8	−3.4 ± 0.3	−1.2 ± 0.2	0.7 ± 0.0	−2.9 ± 0.3	−7.25 ± 0.25
(d) i18c-l78c BF	−4.2	−3.2 ± 0.1	−1.8 ± 0.1	0.0 ± 0.0	−1.4 ± 0.0	−5.96 ± 0.14
(e) t75c-d94c FG	−3.2	−1.9 ± 0.3	−1.0 ± 0.3	−0.8 ± 0.0	−0.1 ± 0.1	−8.73 ± 0.29
(f) t79c-d90c FG	−2.7	−2.4 ± 0.2	−1.1 ± 0.1	−1.7 ± 0.0	0.4 ± 0.2	−7.58 ± 0.07
(g) y81c-r87c FG	−2.1	0.9 ± 0.4	−0.3 ± 0.3	−0.5 ± 0.3	1.7 ± 0.1	−3.86 ± 0.21
(h) e29c-k43c CC′	−2.9	−2.4 ± 0.4	−0.2 ± 0.2	−1.4 ± 0.0	−0.7 ± 0.3	−8.73 ± 0.27
(i) r31c-e41c CC′	−2.6	−1.4 ± 0.2	0.4 ± 0.2	−1.8 ± 0.0	−0.0 ± 0.1	−7.36 ± 0.16
(j) e33c-l38c CC′	−1.9	−1.8 ± 0.4	0.5 ± 0.2	−0.7 ± 0.3	2.0 ± 0.1	−4.2 ± 0.21
(k) a40c-l50c C′C″	−2.6	−4.2 ± 0.4	0.7 ± 0.3	−0.8 ± 0.1	−4.1 ± 0.0	−8.72 ± 0.06
(l) l58c-d62c DE	−1.6	1.5 ± 0.4	−1.3 ± 0.4	0.5 ± 0.1	2.3 ± 0.1	−5.94 ± 0.31
(m) i57c-l63c DE	−2.1	−4.5 ± 0.2	−0.8 ± 0.2	−1.5 ± 0.1	−2.2 ± 0.1	−11.13 ± 0.15

[a] Shown are the effects of introducing a cross-bridge on the free energy changes of defined steps in the folding reaction; all values are derived from the parameters in Table 1 and are given in kcal/mol. Bridge positions are described as in Table 1. The $\Delta\Delta G_{(U\to F)}$ values represent the energetic effect of the bridge on the free energy change for the overall folding reaction as predicted from the entropic relationship: $\Delta\Delta G = -T\Delta\Delta S_{conf} = -T(-2.1 - [1.5\,R \ln n])$ cal/mol, and as measured experimentally (i.e., $\Delta G_{ox} - \Delta G_{red}$). The value $\Delta\Delta G_{(U\to I)}$ represents the effect on the initial rapid transition, and $\Delta\Delta G_{(I\to t)}$ and $\Delta\Delta G_{(t\to F)}$ show their effects on reaching the transition state and relaxing to the folded ground state; they are calculated from the relationships: $\Delta\Delta G_{(I\to t)} = -RT \ln [k_{f(ox)}/k_{f(red)}]$ and $\Delta\Delta G_{(t\to F)} = -RT \ln [k_{u(red)}/k_{u(ox)}]$, respectively. In the last column, absolute stabilities of the oxidized proteins are shown for comparison with the wild-type where $\Delta G_{(U\to F)}$ is −6.2 kcal/mol. Quoted errors are standard errors given to two significant figures and were calculated from fits to the kinetic plots using the standard statistical formula.

reducing its conformational entropy and raising its free energy (*19*); hence, the folded state becomes relatively more favorable. For simplicity, this constraining effect is considered absent in the folded state and can be approximated by the equation: $\Delta\Delta G = -T\Delta\Delta S_{conf} = -T(-2.1 - [1.5R \ln n])$ cal/mol (where n is the number of residues that span the cross-link) which is derived from a statistical mechanics treatment of polymers (*20*). In seven cases, the observed enhancement of stability is close to that predicted, showing both that the empirically derived constant (−2.1) in the above equation is plausible (see Table 2) and that, by and large, the bridges have been well chosen so as not to strain the folded state. For five cases, the observed value is smaller than that predicted. There are two explanations for this. Either the disulfide bridge introduces strain into the folded state, thus raising its enthalpy, or it renders the folded state more rigid, so lowering its entropy. Whichever is the truth, the data do not distinguish between the possibilities.

Only in two cases do the measured increases in stability exceed those predicted from chain entropy. It is noteworthy that in both of these mutants nonpolar residues are replaced by cysteines and, commensurately, the neighboring environment is less polar than for most of the mutation sites. In these circumstances, the more hydrophobic nature of the cystine group, by comparison with two cysteines, will slightly enhance the stability of the oxidized protein over and above the chain-entropy effect of the covalent bridge.

Bridge Position and Differential Effects on the Pathway. More interestingly, the results show that different bridge positions have distinctly different kinetic consequences when the reaction is broken down into its component steps. Figure 3 shows the effects of the cross-links at each stage in the pathway for each mutant and is a graphic representation of the data given in Table 2. Bridges in the nucleus of the molecule formed by the B, C, E, and F strands (see Figures 1 and 4) have their major effect on the initial, rapid step; i.e., they stabilize the transient intermediate. Bridges in the DE and FG hairpins also promote the formation of the intermediate and, in addition, lower the major transition-
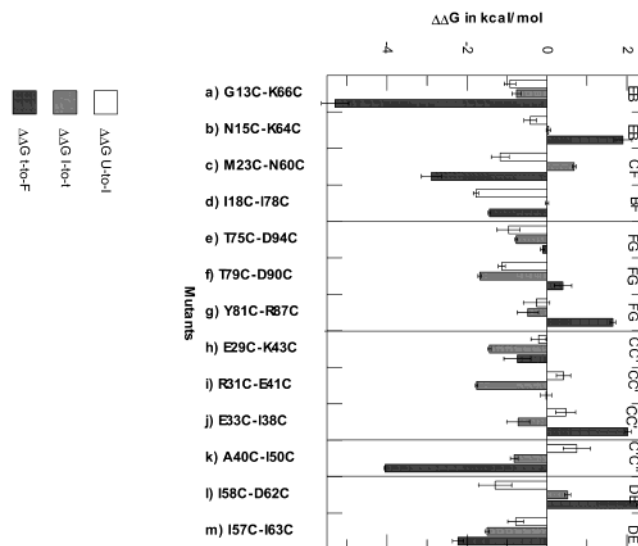

FIGURE 3: Effects of bridges on the free energy of defined steps in the folding pathway. The data derived from the kinetic fits (see Table 1) are represented as a bar chart. The equilibrium and rate constants are converted to changes in free energy; for any transition, $\Delta\Delta G = -RT \ln (k_{ox}/k_{red})$. The bridge positions are indicated by the residue numbers substituted for cysteines in each mutant on the x-axis and by the strands linked above the data.

state barrier; i.e., they enhance the folding rate. Bridges in the CC′ hairpin have little effect on the intermediate but reduce the barrier height for the slow, second step. The C′C″ bridge has no influence in the folding direction but stabilizes the folded state by raising the barrier to unfolding.

For four of the mutants (b, g, j, and l), the introduction of a bridge in the oxidized molecule leads to conformational strain. This effect is apparent at the last stage of folding (rate-limiting transition state to folded state), when side-chains become tightly packed, and is manifested as a positive $\Delta\Delta G$ for the t-to-F transition i.e., an increase in the rate of unfolding.

The order of strand-pairings as CD2 folds, deduced from the data described above, is summarized in graphical form in Figure 4.
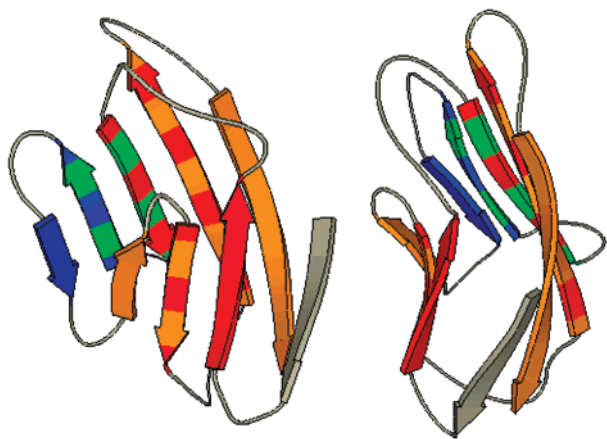
FIGURE 4: Effect of disulfide bridges on folding. The molecule is shown from two angles and color-coded to represent the accumulation of strand-pairings through the folding pathway: Strands containing red are predominantly paired in the initial U-to-I transition, i.e., CF, BE, and BF. Those containing orange (FG and DE) are paired in the U-to-I transition, and the interaction strengthens in the I-to-t transition. Those containing green (CC′) are not paired until the I-to-t transition, and those containing blue (C′C″) only pair after the rate-limiting transition state, so have no influence in the folding direction.

Comparison with Hydrogen Exchange Data. These findings can be compared to measurements of hydrogen exchange protection which showed that the amides most highly protected in the intermediate were in the B, C, E, F, and G strands (*9*). The problem of interpreting these previously determined protection patterns is that the exchange rates of amide groups could be reduced either by burial in the core or by hydrogen bonding to non-native partners. However, the strong correspondence between topological mapping by the bridging method described here and the protection data acquired previously (*9*, *12*) shows that the complex topology of the immunoglobulin fold is established in the very first, sub-millisecond step. The fact that interactions between β-strands that both are sequence-distant (BE, CF) and are in local hairpins (DE, FG) are established in this first step suggests that this early formation of structure is global and concerted. This conclusion is supported by recent experiments on small, single-domain proteins that have kinetically resolved this step and have shown it to have the properties of a first-order transition with a distinct energy barrier separating the unfolded state from the intermediate (*21*, *22*). Furthermore, it has been shown that the fixed, quasi-crystalline interactions between side-chains which are characteristic of the native state are absent in these early folding intermediates (*23*). Taken together, these findings prompt the conclusion that the complex topology of protein folds can be specified without the need for detailed, close-packed side-chain interactions. In turn, this implies that rather crude representations of amino acid residues may suffice in computational models which seek to relate sequence to fold in proteins (*24–26*).

Comparison with Other Studies. Prior to the experiments described here, there have been no extensive studies of the position-dependent effects of cross-links on folding kinetics, but the general principle has been tested in experiments on Barnase. Two bridges were introduced; one in a region which according to phi analysis becomes structured early in the folding pathway and one which is only structured after the

rate-limiting transition state has been surmounted (*15*). Reassuringly, the former bridge led to a more stable intermediate and transition state while the latter affected only the rate of unfolding.

Recently, using a lattice Monte Carlo model, Abkevich and Shakhnovich (*7*) have simulated the effects of introducing disulfide bonds into different regions of a protein, with respect to both the overall molecular stability and the rate of folding. In agreement with the data we describe here, they found a clear correlation between overall stability and the distance along the chain between the cysteine residues. However, there is little agreement between experiment and model with respect to the kinetics of the system. In the Monte Carlo model, the introduction of a bridge at best leads to a 2-fold enhancement of folding rate and in most cases results in a severe deceleration of folding, by up to 3 orders of magnitude. In contrast, we find that the cross-link either has little effect on the folding rate or has a substantial accelerating effect. Only in two cases do we find a small but measurable reduction in folding rate (2–3-fold), whereas in eight cases there is a significant increase. A total of 4 out of the 13 bridge mutants show an enhancement in folding rate greater than a factor of 10.

In the Monte Carlo study, the authors suggest that the very large reductions in folding rate seen in the model may result from constraints on the movement of the simulated polypeptide chain on the cubic lattice. This makes the model system stiffer and thus slower to explore conformational space. It appears that in the real system the greater degree of conformational freedom with respect to torsional and flexing motions of the polypeptide chain enables it to explore conformations in a less restricted way, thus avoiding the problems of stiffness.

Contact Order and Folding Rates. There has been much recent discussion of the factors that affect the rate of protein folding, particularly with respect to the topology and contact order of the molecule (*17*, *18*). Contact order is a measure of the average sequence spacing of interresidue contacts in the native state and, as such, is related to fold architecture. For example, a protein composed of α-helices is dominated by local intrahelical contacts and will have a small contact order compared with a β-sheet protein such as CD2 where a large proportion of native contacts are between residues well separated in sequence. It is argued that the former structure should fold more rapidly owing to the higher probability of forming the native interactions between side-chains. The insertion of a disulfide bond into a protein will also influence contact order because the bridge will effectively reduce the overall sequence spacing of the native contacts (see legend to Figure 5). Indeed, it can be argued that this system provides a good test of such ideas because contact order is being changed with a negligible change in sequence. However, it should be noted that the contact-order analysis has been developed for two-state systems, i.e., those in which there are no populated intermediates. In these cases the reaction in the folding direction can be described by the surmounting of a single transition state barrier. The reaction we deal with in this study is less straightforward in that folding proceeds through three states. Hence, there are two distinct phases in the process, either or both of which could be influenced by the change in contact order. By definition, the contact order is argued to be related to the rates of
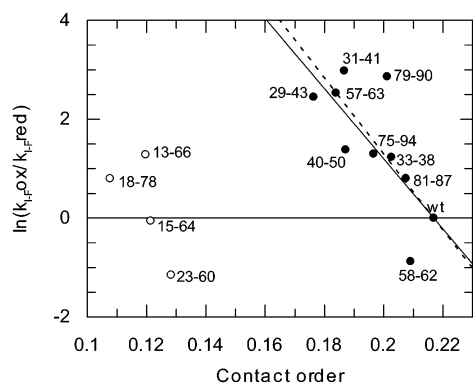
FIGURE 5: Relationship between topology and folding rate. A plot of Contact orders against the natural log of change in folding rate [ln ($k_{I-F(ox)}/k_{I-F(red)}$)] (the comparison is made relative to reduced protein sample as this reports the change in folding kinetics due only to the formation of the bridge). Contact order is calculated by CO = $(\Sigma^N \cdot \Delta S_{i,j})/L \cdot N$ (*17*), where N is the total number of contacts, $\Delta S_{i,j}$ is the sequence separation, in residues, between contacting residues i and j, and L is the total number of residues in the protein. The insertion of a cross-link reduces the contact order by covalently connecting segments of the chain and thus shortening the through-bond sequence separation of residues that make contact in the folded state. The contact order for the cross-bridged molecules was calculated by summing the sequence separations as determined by the smallest number of covalently linked residues between contacting side-chains in the folded state. A contact was deemed to have formed between two residues when their closest side-chain heavy atoms were within 5 Å of each other. The continuous line represents the fit to our own data except for the four proteins where the effect of the bridge is chiefly on the U-I transition; the short dashes represent the data of Baker and colleagues (*17*) (see main text). The two linear fits are shown superimposed upon the data. The black closed circles represent those hairpins that are local in sequence, while open white circles represent bridges that are distant in sequence.

folding, and in this study we can only measure the rate constant for crossing the major transition state, i.e., the I-to-F transition. Hence, in Figure 5 we show the change in this rate constant plotted against the change in contact order occasioned by the insertion of the given bridge. It is striking that the data fall into two distinct categories, and at first sight the correlation between contact order and the logarithm of the rate constant (a measure of the change in barrier height) appears to be scrambled. However, on further inspection the left-hand group of points all represent data for long-range bridges. As described earlier, these bridges have their major influence on the rapid first step in folding, the rate of which is immeasurably fast using a stopped-flow apparatus. Because their effect is predominantly on this step, the influence on the second step is much smaller than expected by the contact-order argument. By contrast, for those bridges which have their major effect on the energy of the transition state, i.e., upon k_{I-F} (shown as closed circles in Figure 5), there is a more plausible relationship between contact order and folding rate. Super-imposed upon the data in Figure 5 is the slope value determined by Baker and colleagues (*17*) for naturally occurring proteins (−77). For those bridges which have their primary effect upon the rate of folding (closed circles), the correspondence appears reasonable. The above discussion highlights the most surprising finding of this study, that the slow, rate-limiting process in the folding of this complex β-sheet structure is not the formation of sequence-distant strand−strand interactions; these are rapid, early events.

Rather, the slow step represents the formation of more localized, hairpin structures once the complex elements of topology are established. The observation that long-range contacts and therefore molecular order can be established in early, non-rate-limiting processes precludes the application of simple contact-order/rate analysis to such three-state systems.

Implications for Protein Engineering. It is worth considering implications these observations have for the rational design of more robust proteins. Previous, less systematic studies have concluded that the effect of disulfide bridges on protein stability is unpredictable (*27, 28*). However, the results of this work show that for β-sheet structures, surface disulfide bonds can readily be inserted at non-hydrogen-bonded sites between β-strands on the protein surface to produce hyper-stable native states with an added stability equal to or greater than that predicted by the statistical entropy rule. In their oxidized states, 8 out of 13 are more stable than the wild-type molecule with 2 being more than 3500-fold and 5 being more than 50-fold more stable with respect to the equilibrium constant for folding. It is worthy of note that 45% of all proteins whose structures have been determined have exposed β-sheets that could be stabilized by surface disulfide engineering.

REFERENCES

1. Matouschek, A., Kellis, I. T., Serrano, I., and Fersht, A. R. (1989) Mapping the transition-state and pathway of protein folding by protein engineering. *Nature 340*, 122−126.
2. Fersht, A. R. (1997) Nucleation mechanisms in protein folding. *Curr. Opin. Struct. Biol. 7*, 3−9.
3. Ellis, R. J., Dobson, C., and Hartl, U. (1998) Sequence does specify protein conformation. *Trends Biochem. Sci. 23*, 468−468.
4. Englander, S. W., and Mayne, L. (1993) Protein folding studied using hydrogen-exchange labeling and 2-dimensional NMR. *Annu. Rev. Biophys. Biomol. Struct. 21*, 243−265.
5. Clarke, J., and Fersht, A. R. (1993) Engineered disulfide bonds as probes of the folding pathway of barnase−increasing the stability of proteins against the rate of denaturation. *Biochemistry 32*, 4322−4329.
6. Yokota and Segawa, S. (2000) The transition state in the folding−unfolding reaction of four species of three-disulfide variant of hen lysozyme: the role of each disulfide bridge. *J. Mol. Biol. 295*, 1275−1288.
7. Abkevich, V. I., and Shakhnovich, E. I. (2000) What can disulfide bonds tell us about protein energetics, function and folding: simulations and bioinformatics analysis. *J. Mol. Biol. 300*, 975−985.
8. Grantcharova, V. P., Riddle, D. S., and Baker, D. (2000) Long-range order in the src SH3-folding transition state. *Proc. Natl. Acad. Sci. U.S.A. 97*, 7084−7089.
9. Parker, M. J., Dempsey, C. E., Hosszu, L. P., Waltho, J. P., and Clarke, A. R. (1998) Topology, sequence evolution and folding dynamics of an immunoglobulin domain. *Nat. Struct. Biol. 5*, 194−198.
10. Parker, M. J., Lorch, M., Sessions, R. B., and Clarke (1998) Thermodynamic properties of transient intermediates and transition states in the folding of two contrasting protein structures. *Biochemistry 37*, 2538−2545.
11. Lorch, M., Mason, J. M., Clarke, A. R., and Parker, M. J. (1999) Effects of core mutations on the folding of a beta-sheet protein: implications for backbone organization in the I-state. *Biochemistry 38*, 1377−1385.

12. Parker, M. J., Dempsey, C. E., Lorch, M., and Clarke, A. R. (1997) Acquisition of native beta-strand topology during the rapid collapse phase of protein folding. *Biochemistry 36*, 13396−13405.

13. Hutchinson, E. G., Sessions, R. B., Thornton, J. M., and Woolfson (1998) Determinants of strand register in antiparallel beta-sheets of proteins. *Protein Sci. 7*, 2287−2300.

14. Wouters, M. A., and Curmi, P. M. G. (1995) An analysis of side-chain interactions and pair correlations within antiparallel beta-sheets—the differences between backbone hydrogen-bonded and non-hydrogen-bonded residue pairs. *Proteins: Struct., Funct., Genet. 22*, 119−131.

15. Matsumura, M., Signor, G., and Matthews, B. (1989) Substantial increase of protein stability by multiple disulfide bonds. *Nature 342*, 291−293.

16. Schonbrunner, N., Pappenburger, G., Scharf, M., Engels, J., and Kiefhaber, T. (1997) Effect of preformed correct tertiary interactions on rapid two state Tendamistat folding: evidence for hairpins as initiation sites for *β*-sheet formation. *Biochemistry 36*, 9057−9065.

17. Plaxco, K. W., Simons, K. T., and Baker, D. (1998) Contact order, transition state placement and the refolding rates of single domain proteins. *J. Mol. Biol. 277*, 985−994.

18. Fersht, A. R. (2000) Transition-state structure as a unifying basis in protein-folding mechanisms: contact order, chain topology, stability, and the extended nucleus mechanism. *Proc. Natl. Acad. Sci. U.S.A. 97*, 1525−1529.

19. Flory, P. J. (1956) Theory of elastic mechanisms in fibrous proteins. *J. Am. Chem. Soc. 78*, 5222−5235.

20. Pace, C. N., Grimsley, G. R., Thompson, J. A., and Barnett, B. J. (1988) Conformational stability and activity of ribonuclease-T1 with zero, one, and two intact disulfide bonds. *J. Biol. Chem. 263*, 11820−11825.

21. Capaldi, A. P., Shastry, M. C. R., Kleanthous, C., Roder, H., and Radford, S. E. (2001) Ultrarapid mixing experiments reveal that Im7 folds via an on-pathway intermediate. *Nat. Struct. Biol. 8*, 68−72.

22. Shastry, M. C. R., and Roder, H. (1998) Evidence for barrier-limited protein folding kinetics on the microsecond time scale. *Nat. Struct. Biol. 5*, 385−392.

23. Ptitsyn, O. B. (1992) in *Protein Folding (1992): The Molten Globule as a General Kinetic Intermediate, Nativelike Features of Structural Organization* (Creighton, T., Ed.) pp 286−290, W. H. Freeman & Co, New York.

24. Sippl, M. J. (1995) Knowledge-based potentials for proteins. *Curr. Opin. Struct. Biol. 5*, 229−235.

25. Miyazawa, S., and Jernigan, R. L. (1996) Residue−residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading. *J. Mol. Biol. 256*, 623−644.

26. Gibbs, N., Sessions, R. B., and Clarke, A. R. (2001) Ab initio protein structure prediction using physio-chemical potentials. *Proteins: Struct., Funct., Genet. 43*, 186−202.

27. Mansfeld and Eijsink, V. G. (1997) Extreme stabilization of a thermolysin-like protease by an engineered disulfide bond. *J. Biol. Chem. 272*, 11152−11156.

28. Mitchinson, C., and Wells, J. A. (1989) Protein engineering of disulfide bonds in subtilisin Bpn′. *J. Biochem. 28*, 4807−4815.

29. Driscoll, P. C., Cyster, J. G., Campbell, I. D., and Williams (1991) Structure of domain-1 of rat lymphocyte-T CD2 antigen. *Nature 353*, 762−765.

30. Parker, M. J., and Clarke, A. R. (1997) Amide backbone and water-related H/D isotope effects on the dynamics of a protein folding reaction. *Biochemistry 36*, 5786−5794.

31. Creighton, T. E., Ed. (1990) in *Protein Structure: A Practical Approach* Chapter 7, IRL Press, Oxford.

32. Parker, M. J., Spencer, J., and Clarke, A. R. (1995) An integrated kinetic-analysis of intermediates and transition-states in protein-folding reactions. *J. Mol. Biol. 253*, 771−786.