Probing minimal independent folding units in dihydrofolate reductase

ARTICLE *in* PROTEIN SCIENCE · SEPTEMBER 1997

Impact Factor: 2.85 · DOI: 10.1002/pro.5560060909 · Source: PubMed

3 AUTHORS, INCLUDING:

Charles Robert Matthews
University of Massachusetts Medical School
146 PUBLICATIONS **5,902** CITATIONS

Protein Science (1997), 6:1885–1892. Cambridge University Press. Printed in the USA.

Probing minimal independent folding units in dihydrofolate reductase by molecular dissection

COLIN V. GEGG,[1] KATHERINE E. BOWERS, AND C. ROBERT MATTHEWS

Department of Chemistry and Center for Biomolecular Structure and Function, The Pennsylvania State University, University Park, Pennsylvania 16802

(RECEIVED March 12, 1997; ACCEPTED May 9, 1997)

Abstract

Molecular dissection was employed to identify minimal independent folding units in dihydrofolate reductase (DHFR) from *Escherichia coli*. Eight overlapping fragments of DHFR, spanning the entire sequence and ranging in size from 36 to 123 amino acids, were constructed by chemical cleavage. These fragments were designed to examine the effect of tethering multiple elements of secondary structure on folding and to test if the secondary structural domains represent autonomous folding units. CD and fluorescence spectroscopy demonstrated that six fragments containing up to a total of seven α-helices or β-strands and, in three cases, the adenine binding domain (residues 37–86), are largely disordered. A stoichiometric mixture of the two fragments comprising the large discontinuous domain, 1–36 and 87–159, also showed no evidence for folding beyond that observed for the isolated fragments. A fragment containing residues 1–107 appears to have secondary and tertiary structure; however, spontaneous self-association made it impossible to determine if this structure solely reflects the behavior of the monomeric form. In contrast, a monomeric fragment spanning residues 37–159 possesses significant secondary and tertiary structure. The urea-induced unfolding of fragment 37–159 in the presence of 0.5 M ammonium sulfate was found to be a well-defined, two-state process. The observation that fragment 37–159 can adopt a stable native fold with unique, aromatic side-chain packing is quite striking because residues 1–36 form an integral part of the structural core of the full-length protein.

Keywords: autonomous folding units; CD; chemical cleavage; fluorescence; protein folding; protein fragments

Protein folding reactions at equilibrium usually involve only the native and the unfolded conformations at appreciable concentrations (Matthews, 1993). Although this simple two-state behavior is advantageous for quantitative measurements of their thermodynamic properties, it has been a major impediment to the solution of the protein folding problem. Detailed structural, thermodynamic, and dynamic information on partially folded forms that might be expected to provide insight into the folding mechanism is often not available.

One approach toward using high-resolution equilibrium methods to study partially folded proteins is to alter the solution conditions so as to destabilize the native conformation without simultaneously stabilizing the fully unfolded form. Acidic pH or high temperatures have been used routinely to populate molten globule species that have substantial secondary structure but little or no well-defined tertiary structure (Baldwin, 1991; Skolnick et al., 1993; Dobson, 1994). In some cases (Dabora et al., 1996; Kuwajima, 1996; Raschke & Marqusee, 1997), molten globules have been shown to be similar to early kinetic folding intermediates. Measurements of hydrogen exchange rates of backbone amide hydrogens at low denaturant concentrations (Bai & Englander, 1996) can reveal very small populations of partially folded forms in equilibrium with the native conformation. Some of these minor conformers contain secondary structure that resembles that found in early kinetic intermediates (Bai & Englander, 1996; Chamberlain et al., 1996).

Another approach toward breaking down the cooperativity of the folding reaction and populating alternatively folded conform-

Reprint requests to: Robert Matthews, Department of Chemistry and Center for Biomolecular Structure and Function, The Pennsylvania State University, University Park, Pennsylvania 16802; e-mail: crm@psu.edu.

[1]Present address: AMGEN Inc., Amgen Center, Thousand Oaks, California 91320.

Abbreviations: ABD, adenine binding domain; Am_2SO_4, ammonium sulfate; AS DHFR, Cys 85 Ala/Cys 152 Ser, cysteine-free double mutant of dihydrofolate reductase; CMW, center-of-mass wavelength; DHFR, dihydrofolate reductase; DTE, dithioerythritol; IMAC, immobilized metal affinity chromatography; Ni^{2+}-IDA, nickel chelated iminodiacetic acid resin; NADH, nicotinamide adenine dinucleotide, reduced form; NADPH, nicotinamide adenine dinucleotide phosphate, reduced form; NADP+, nicotinamide adenine dinucleotide phosphate, oxidized form; NTCB, 2-nitro-5-thiocyanobenzoic acid.

ers is to modify the protein, e.g., by removing an essential ligand or by protein engineering (Shortle et al., 1996). For example, removal of the heme from myoglobin (Eliezer & Wright, 1996; Lecomte et al., 1996) or cytochrome *b*5 (Falzone et al., 1996) results in species that lose structure in the heme pocket, but retain a native-like fold in other regions. Site-specific amino acid replacements in staphylococcal nuclease (Shortle & Meeker, 1986) and tryptophan repressor (Mann et al., 1993) uncouple the disruption of secondary and tertiary structure by denaturants, consistent with a significant population of a partially folded form(s). The removal of 18 amino acids from the N and C termini of staphylococcal nuclease, in combination with several point mutations, yields a partially folded form that remains native-like in the β-sheet region, but has substantially less structure in the helix and loop region (Alexandrescu et al., 1994).

A logical extension of manipulating the amino acid sequence of a protein to reduce the cooperativity of its folding reaction is to produce fragments of various sizes and determine whether the independent segments are capable of folding. Although peptides representing individual elements of secondary structure often do not fold significantly (Dyson et al., 1992a, 1992b), larger fragments containing multiple elements or even entire structural domains offer the possibility of mutual stabilization. For example, linking pairs of individually unstructured peptides from bovine pancreatic trypsin inhibitor with a disulfide bond results in the formation of native-like secondary and tertiary structure (Oas & Kim, 1988; Staley & Kim, 1994). Some, but not all, proteolytic fragments of tryptophan repressor and cytochrome *c* fold spontaneously, revealing autonomous folding subdomains (Tasayco & Carey, 1992; Wu et al., 1993, 1994). The two structural domains in yeast phosphoglycerate kinase have been found to fold independently (Missiakas et al., 1990); similar behavior is observed in other multidomain proteins (Jaenicke, 1991). Finally, Peng and Kim (1994) have found that fusing the two discontinuous segments of the α-helical domain in α-lactalbumin resulted in a molten globule-like conformation with a native-like fold. These studies demonstrate that useful insights into the distribution of folding information in the primary structure of a protein can be obtained by examining the behavior of fragments.

Building upon these successful applications, the fragmentation approach was applied to *Escherichia coli* dihydrofolate reductase. DHFR is a 159-amino acid, monomeric enzyme that catalyzes the reduction of 7,8-dihydrofolate to 5,6,7,8-tetrahydrofolate using NADPH as the reducing cofactor. The tertiary structure of this enzyme is a doubly wound, parallel α/β-sheet, comprised of four α-helices and eight β-strands. This designation places DHFR in the largest subgroup of the most common class of protein structures (Richardson, 1981), offering the possibility that its folding behavior might be representative of a large number of proteins.

The crystal structure of the DHFR-folate-$NADP^+$ ternary complex, solved by Bystroff et al. (1990), shows the presence of two structural domains (Fig. 1). Residues 38–88 define an adenine binding domain (ABD) that is similar to those found in many other NADH or NADPH-dependent enzymes. The sequence of the ABD is flanked by residues 1–37 and 89–159, which define a larger, discontinuous domain. The two-domain structure of DHFR provided the starting point for its molecular dissection and led to the preparation and isolation of a set of eight fragments that incorporate these domains and span the entire sequence. The present study reports the characterization of these fragments in terms of their propensities to fold independently and to adopt secondary and/or tertiary structure.

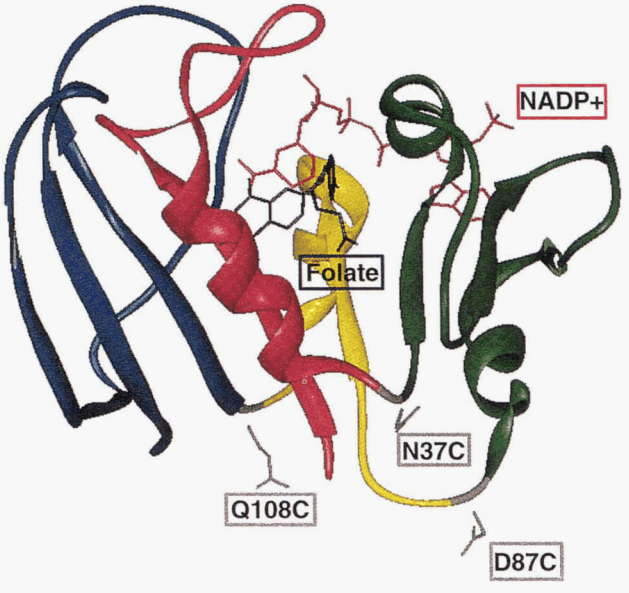

Fig. 1. Structural representation of the ternary complex of dihydrofolate reductase from *E. coli* with $NADP^+$ and folate. Coordinates are from the Brookhaven Protein Data Bank, file 7dfr. Fragments are indicated as follows: 1–36 in magenta; 37–86 in green; 87–107 in yellow; and 108–159 in blue.

Results

Fragment design, construction, and purification

On the basis of the X-ray structure of DHFR (Fig. 1), a set of eight fragments were selected to test the effect of tethering sequential elements of secondary structure and of constructing independent structural domains. These fragments (and the elements of secondary structure spanned in the fully folded protein) include residues 1–36 (α_1; β_1), 1–86 (α_1-α_3; β_1-β_4), 1–107 (α_1-α_4; β_1-β_5), 37–86 (α_2-α_3; β_2-β_4), 37–107 (α_2-α_4; β_2-β_5), 37–159 (α_2-α_4; β_2-β_8), 87–159 (α_4; β_5-β_8), and 108–159 (β_6-β_8). The small discrepancies between these fragments and the boundaries of the domains identified in the X-ray structure reflect the approach adopted to prepare these fragments (see below).

Fragments 1–36 and 87–159 were used to probe the propensity of the two segments of the large, discontinuous domain to fold; a stoichiometric mixture of these two fragments tested the folding of this domain. Fragment 37–86 examined the folding of the ABD, a subdomain originally proposed by Bystroff et al. (1990). Fragment 37–107 tested the folding of an extended ABD whose additional 29 residues introduce several adenosine contact points, a β-strand, and an α-helix whose dipole moment has been proposed to provide a substantial portion of the cofactor binding energy (Bystroff et al., 1990). These additional elements of secondary structure dock directly on the ABD (Fig. 1). Recent crystallographic studies (Sawaya & Kraut, 1997) indicate that this extended fragment may be a more appropriate model of the adenosine binding subdomain given the observation that it behaves as a rigid body in the catalytic mechanism of dihydrofolate reductase. Fragments 1–86 and 37–159 tested the effect of fusing the ABD to either the amino or carboxy region of the larger domain, whereas 108–159 provided an opportunity to probe the behavior of a structurally self-contained portion of the larger domain.

The fragments corresponding to these segments were prepared by chemical cleavage and affinity chromatography (Gegg et al., 1996). Previous studies on DHFR (Iwakura et al., 1995) have shown that the two naturally occurring cysteine residues at positions 85 and 152 can be replaced with alanine and serine, respectively, without significant effects on stability or function. This double mutant, AS-DHFR, serves as an appropriate host for cysteine replacements that are cyanylated readily by NTCB (Jacobson et al., 1973). Peptide bond cleavage occurs under mildly alkaline conditions and results in a normal carboxy terminus for the N-terminal fragment ending just before the cysteine residue. The C-terminal fragment is capped with an iminothiozolidinyl ring that is neutral and slightly larger than a cysteine side chain.

The desired fragments were constructed by introducing cysteine residues either singly or pairwise at positions 37, 87, or 108 in AS-DHFR. Positions 37 and 87 were chosen to be as close as possible to the domain boundaries identified in the X-ray structure and yet be solvent accessible so as to cause a minimal perturbation of the native conformation by the mutations. As described above, position 108 was selected to examine the behavior of an extended ABD fragment.

The purification of the fragments was simplified greatly by incorporating polyhistidine tags at either or both termini of the full-length protein and performing affinity chromatography on a Ni^{2+}-iminodiacetic Sepharose resin. Polyhistidine tags were placed so that none of the purified fragments retained a tag. These procedures and verification of the fragment products by mass spectroscopy are described in detail in Gegg et al. (1996).

Spectroscopic survey of DHFR fragments

The far-UV CD spectra of the eight DHFR fragments and native, full-length AS-DHFR are shown in Figure 2. The spectrum of full-length AS-DHFR has a minimum at ≈218 nm and a maximum at ≈195 nm that reflects its α/β-sheet motif. The spectra of 1–36, 87–159, and 108–159, in contrast, display small shoulders of negative mean residue ellipticity at ≈220 nm and substantial minima

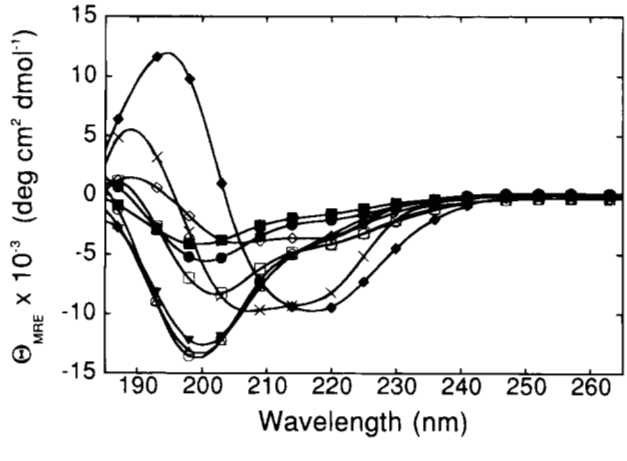

Fig. 2. Far-UV CD spectra of isolated fragments from dihydrofolate reductase, taken in 10 mM potassium phosphate, pH 7.8, 0.2 mM potassium EDTA, and 1 mM β-mercaptoethanol at 15 °C. Full-length AS DHFR (♦), fragments 1–36 (○), 1–86 (□), 1–107 (◇), 37–86 (●), 37–107 (■), 37–159 (×), 87–159 (▼), and 108–159 (△). Spectra are plotted as mean residue ellipticity as a function of wavelength. Fragment concentrations range from 4.1 to 18.8 μM.

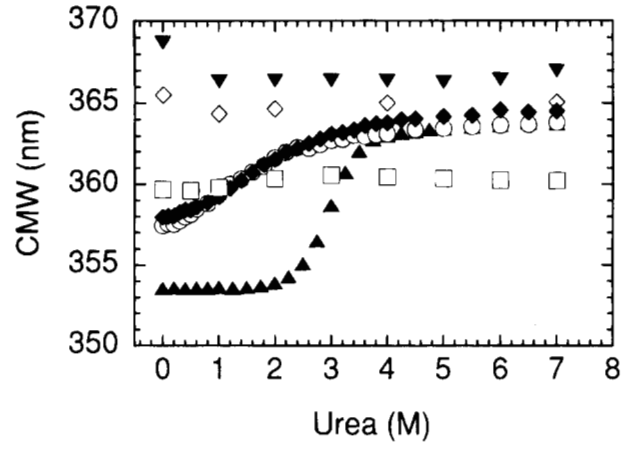

Fig. 3. Urea denaturation of fragments of DHFR monitored by tryptophan fluorescence, taken in 10 mM potassium phosphate, pH 7.8, 0.2 mM potassium EDTA, and 1 mM β-mercaptoethanol, 15 °C. Data are plotted as the CMW as a function of urea. Full-length AS DHFR (▲), fragments 1–36 (◇), 1–107 (○), 37–107 (▼), 37–159 (♦), and 108–159 (□). Fragment concentrations range from 1.9 to 9.5 μM.

at ≈200 nm, implying that these fragments have little tendency to fold. A stoichiometric mixture containing 3.8 μM each of fragment 1–36 and fragment 87–159 yields a CD spectrum that is identical to the sum of the spectra of the individual fragments (data not shown). The absence of an enhanced signal demonstrates that these two fragments, comprising the large discontinuous domain, do not spontaneously associate under these conditions.

The CD spectra of fragments 1–86, 37–86, 37–107, and 1–107 also have small shoulders near 220 nm, but differ from the spectra of 1–36, 87–159, and 108–159 in having significantly reduced negative ellipticities for the minima near 200 nm. Fragment 1–107 also displays a small positive ellipticity near 190 nm. However, the tendency of 1–107 to self-associate at micromolar concentrations (C.V. Gegg & C.R. Matthews, unpubl. size exclusion chromatography data) leaves uncertain the spectroscopic properties of the monomeric species. These spectra suggest that the second set of fragments are not entirely disordered and may have a slight tendency to adopt secondary structure. Interestingly, this set differs from the first set by the presence of the ABD, residues 37–87.

Fragment 37–159 is unique in that it has a far-UV CD spectrum consistent with the presence of substantial secondary structure: a significant shoulder at ≈220 nm, a minimum at ≈208 nm, and a strong, positive signal at ≈190 nm. Unlike 1–107, fragment 37–159 is monomeric at concentrations equivalent to those used for CD spectroscopy (size exclusion chromatography data, not shown).

Tertiary structure involving the tryptophan residues in all eight fragments and the full-length protein was probed by fluorescence spectroscopy. AS-DHFR has tryptophans at positions 22, 30, 47, 74, and 132, providing one or more chromophores in each fragment. The fluorescence spectra of fragments 1–36, 1–86, 37–86, 37–107, 87–159, and 108–159 displayed emission maxima near 350 nm (data not shown). Further, the CMW of these fragments are not sensitive to the presence of urea (see, e.g., the behavior of 1–36, 37–107, and 108–159 in Fig. 3). These results are consistent with the full solvent exposure of the tryptophan side chains and their lack of involvement in tertiary structure. As with CD spectroscopy, the fluorescence spectrum of a stoichiometric mixture of fragments 1–36 and 87–159 is identical to the sum of the spectra

of the individual fragments (data not shown). This result provides further support for the conclusion that these fragments do not associate spontaneously under these conditions.

In contrast, the CMW of fragments 1–107 and 37–159 undergo a significant and apparently cooperative red shift in the presence of urea (Fig. 3). For 37–159, these results imply that this fragment partially buries one or more of its three tryptophans in a hydrophobic environment.

The tertiary structure in fragment 37–159 was further examined by near-UV CD spectroscopy to determine if its aromatic side chains have unique packing environments (Fig. 4A). The near-UV CD spectrum of this fragment has a significant minimum at 287 nm and elements of fine structure between 265 and 295 nm, implying that the orientation of one or more of the aromatic side chains are constrained by tertiary structure. The fine structure of 37–159 occurs at similar wavelengths to those observed in AS-DHFR (Fig. 4C); however, the amplitudes and/or signs of the signals are significantly different. These differences could arise from alternative packing environments near the chromophores and/or from the absence of two tryptophans and one phenylalanine in the fragment.

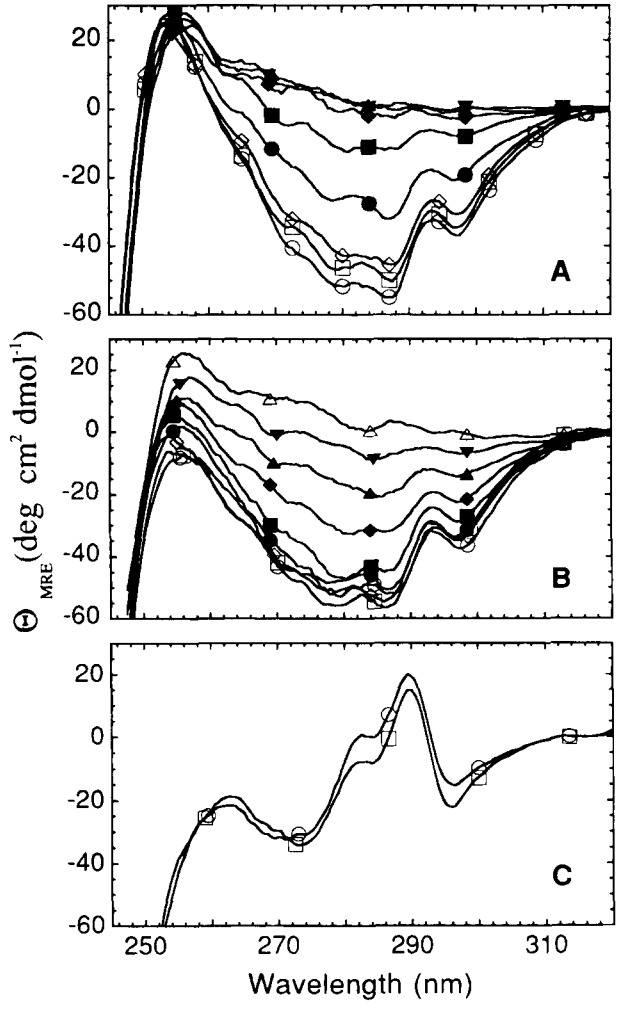

Fig. 4. Urea dependence of the near-UV CD spectra of fragment 37–159. **A:** Fragment spectra at 0 M (○), 0.25 M (□), 0.5 M (◇), 1.0 M (●), 1.5 M (■), 2.0 M (◆), 2.5 M (▲), and 3.0 M (▼) urea. **B:** Fragment spectra in 0.5 M Am_2SO_4 at 0 M (○), 0.5 M (□), 1.0 M (◇), 1.5 M (●), 2.0 M (■), 2.5 M (◆), 3.0 M (▲), 3.5 M (▼), and 6.0 M (△) urea. **(C)** Full-length AS DHFR in the presence (□) and absence (○) of 0.5 M Am_2SO_4. Buffer is 10 mM potassium phosphate, pH 7.8, 0.2 mM potassium EDTA, 1 mM β-mercaptoethanol, 15 °C. The concentration of fragment 37–159 is 7.7 μM. The concentration of AS DHFR is 4.4 μM.

Fragment stability

The urea-induced denaturation of each fragment and AS-DHFR was monitored by intrinsic tryptophan fluorescence and far-UV CD spectroscopy (Figs. 3 and 5, respectively). Fragments 1–36, 37–107, and 108–159 displayed little or no dependence of either signal on the urea concentration and are representative of the behavior of fragments 1–86, 37–86, and 87–159 (data not shown). This response is consistent with the minimal amount of secondary structure evident in their far-UV CD spectra (Fig. 2) and the solvent exposure of the tryptophans in the absence of denaturant.

Only fragment 37–159 demonstrated a cooperative unfolding transition that could be ascribed to a unimolecular reaction (Figs. 3, 5). The far-UV CD signal (Fig. 5) begins to decrease at very modest urea concentrations, demonstrating that the fully folded form has marginal stability. The unfolding of this fragment occurs with a transition region between 0 and 2 M urea, with little or no native baseline. Similar to the far-UV CD data, the unfolding transition monitored by the CMW of the tryptophan emission (Fig. 3) indicates a moderately cooperative process that occurs over the same range in urea concentration as that for the transition detected in the far-UV CD experiment. The decrease in the near-UV CD signal at low urea concentrations (Fig. 4A) parallels the transition detected by fluorescence spectroscopy, demonstrating that the exposure of tryptophans to solvent is accompanied by the loss in tight packing around the aromatic chromophores.

The reliable comparison of folding models and the quantitative analysis of thermodynamic parameters from denaturant-induced unfolding reactions require accurate estimates of the optical properties of both the fully folded and fully unfolded forms. Unfortunately, the marginal stability of 37–159 (Figs. 3, 4, 5) prohibits an accurate assessment of the optical properties of its fully folded conformation. This problem was remedied by adding a stabilizing agent, ammonium sulfate, to the buffer.

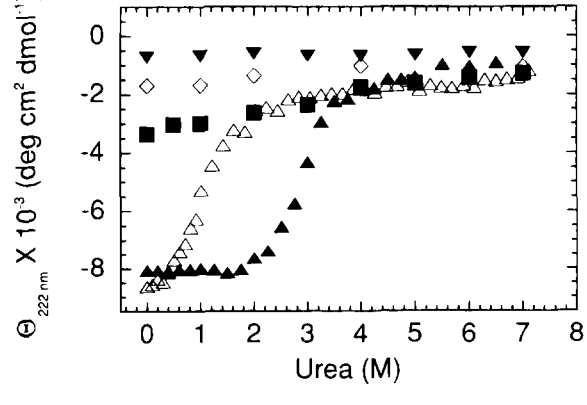

Fig. 5. Urea denaturation of fragments of DHFR monitored by CD, plotted as mean residue ellipticity at 222 nm as a function of urea. Full-length AS DHFR (▲), fragments 1–36 (◇), 37–107 (▼), 37–159 (△), and 108–159 (■). Buffer and conditions are identical to those described in Figure 2. Fragment concentrations range from 1.9 to 9.5 μM.

The effects of 0.5 M Am_2SO_4 on the urea-induced unfolding reaction of 37–159, monitored by far-UV CD (222 nm), near-UV CD (292 nm), and fluorescence spectroscopy (CMW), are shown in Figure 6. These methods illustrate that the unfolding of fragment 37–159 results in a cooperative transition with the fully folded form of this fragment populated up to nearly 1.5 M urea in the presence of ammonium sulfate. This behavior is sufficient to permit a reliable thermodynamic analysis.

The optical transitions shown in Figure 6 can be replotted in terms of apparent fraction of unfolded protein, F_{app}, to allow direct comparisons of the disruption of secondary and tertiary structure (Fig. 7) (Finn et al., 1991). F_{app} plots derived from the far-UV CD, near-UV CD, and tryptophan fluorescence data for fragment 37–159 in 0.5 M Am_2SO_4 are coincident (Fig. 7), indicative of a two-state reaction.

The thermodynamic parameters for 37–159 and those for the full-length protein in the same solvent are shown in Table 1. For the fragment, the free energy of folding in the absence of denaturant, $\Delta G°(H_2O)$, the sensitivity of the transition to denaturant, m, and the midpoint of the transition, C_m, determined by the local fits of the near-UV CD, far-UV CD, and fluorescence data, are within error to those parameters derived from the global fit of the three data sets. A reduced chi-squared value of 1.35 gives statistical evidence for the accuracy of the two-state equilibrium model.

In the presence of 0.5 M Am_2SO_4, 37–159 has a stability of 3.12 ± 0.24 kcal mol^{-1}, an m value of 1.35 ± 0.09 kcal mol^{-1} M^{-1}, and a C_m of 2.31 ± 0.24 M urea. For comparison, the free energy of folding of the full-length protein in the absence of denaturant at pH 7.8, 15 °C, and 0.5 M Am_2SO_4 is 11.00 ± 0.65 kcal mol^{-1}, with an m value of 1.97 ± 0.12 kcal mol^{-1} M^{-1} and a C_m of 5.58 ± 0.48 M urea. The free energy is 75% greater than that without the stabilizing salt, 6.28 ± 0.17 kcal mol^{-1} (Table 1), highlighting the effect of Am_2SO_4.

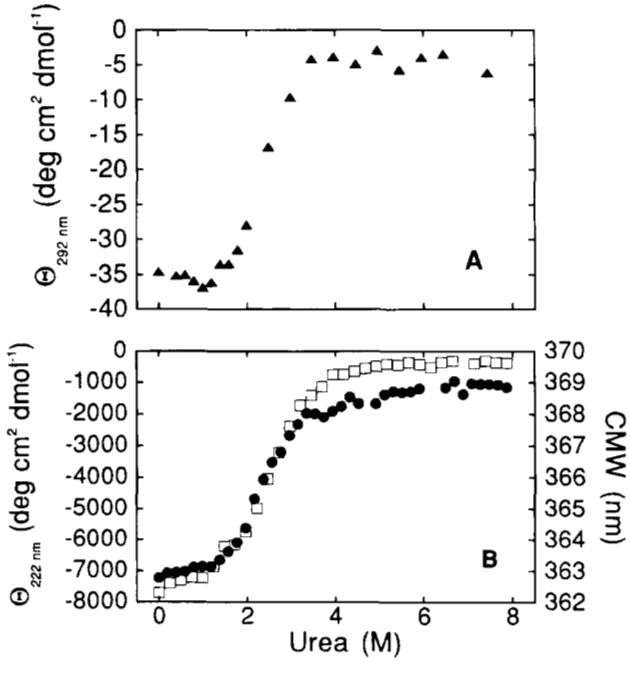

Fig. 6. Optical data for the equilibrium urea titrations of fragment 37–159 in the presence of 0.5 M Am_2SO_4. **A:** As monitored by near-UV CD, presented as mean residue ellipticity at 292 nm (▲). **B:** As monitored by tryptophan fluorescence (□), presented as the CMW, and far-UV CD (●), presented as mean residue ellipticity at 222 nm. Buffer is 10 mM potassium phosphate, pH 7.8, 0.2 mM potassium EDTA, 1 mM β-mercaptoethanol, 15 °C. The concentrations of fragment 37–159 are 7 μM for near-UV CD, 4.1 μM for far-UV CD and 3.6 μM for fluorescence.

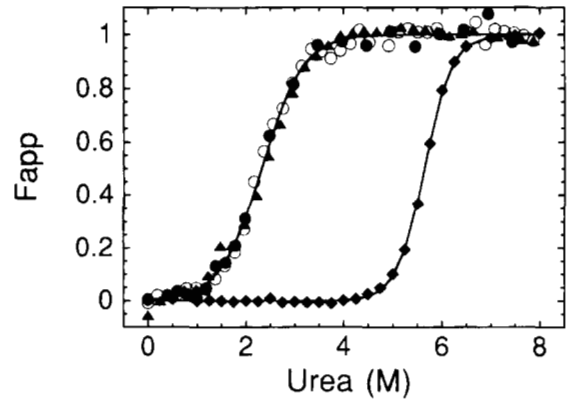

Fig. 7. Normalized equilibrium unfolding data in the presence of 0.5 M Am_2SO_4 of fragment 37–159 monitored by: tryptophan fluorescence (▲), CD at 222 nm (○), and CD at 292 nm (●) with data fitting to a global two-state model. Full-length AS DHFR monitored by tryptophan fluorescence with data fitting to a two-state model (♦).

Discussion

Implications for the folding mechanism of DHFR

The absence of secondary or tertiary structure in fragments 1–36, 87–159, and 108–159 shows that simply tethering up to five α-helices and/or β-strands from AS-DHFR does not increase their propensity to fold. Although this observation does not rule out the framework model (Kim & Baldwin, 1982), it does demonstrate that the formation of secondary structure in the two segments that

Table 1. *Thermodynamic parameters for AS-DHFR and fragment 37–159 unfolding by urea in 0.5 M Am_2SO_4*[a]

Fragment	$\Delta G°(H_2O)$ (kcal/mol)	m (kcal/mol/M)	C_m (M)
AS DHFR			
Far-UV CD	5.52 ± 0.49	1.92 ± 0.17	2.87 ± 0.35
Fluorescence	6.39 ± 0.09	2.16 ± 0.03	2.96 ± 0.06
Global	6.28 ± 0.17	2.13 ± 0.06	2.95 ± 0.12
With 0.5 M Am_2SO_4 AS DHFR			
Far-UV CD	10.68 ± 1.17	1.93 ± 0.22	5.54 ± 0.87
Fluorescence	11.57 ± 0.27	2.05 ± 0.05	5.64 ± 0.19
Global	11.00 ± 0.65	1.97 ± 0.12	5.58 ± 0.48
37–159 DHFR			
Far-UV CD	3.85 ± 0.49	1.72 ± 0.22	2.24 ± 0.40
Fluorescence	3.32 ± 0.25	1.27 ± 0.08	2.61 ± 0.26
Near-UV CD	3.29 ± 0.59	1.44 ± 0.22	2.29 ± 0.54
Global	3.12 ± 0.24	1.35 ± 0.09	2.31 ± 0.24

[a]Conditions: 10 mM potassium phosphate, 0.2 mM K_2EDTA, 1 mM β-mercaptoethanol, pH 7.8, 15 °C.

correspond to the larger, discontinuous domain is not a strong driving force in the folding of DHFR. The absence of secondary and tertiary structure in a 1:1 mixture of 1–36 and 87–159 also suggests that the folding of the larger domain is not an important factor in the mechanism. This conclusion must be tempered by the fact that the attempted reconstruction of the discontinuous domain was performed with two independent fragments. The entropy penalty produced by this approach might be sufficient to destabilize this domain.

In contrast, the far-UV CD spectra of the three fragments that encompass the ABD, 1–86, 37–86, and 37–107, suggest that these fragments contain some secondary structure. Estimation of the secondary structure distribution from these spectra is complicated by possible exciton coupling contributions from tryptophans 47 and 74 observed in the full-length protein (Kuwajima et al., 1991). Although the secondary structures are ambiguous, the urea titration data demonstrate that the ABD, similar to the large, discontinuous domain, does not fold to a thermodynamically stable structure.

Therefore, the chemical cleavage of DHFR into fragments that reflect its domain structure has shown that neither the ABD, an extended version of the ABD that includes another α-helix and β-strand, nor the large, discontinuous domain folds to a significant extent in solution. The lack of a correlation between structural domains and stable, autonomous folding units in AS-DHFR means that these domains cannot act as templates to direct subsequent events in folding, as prescribed by the subdomain model (Oas & Kim, 1988).

The hydrophobic collapse model (Dill, 1985; Dill et al., 1989), however, might explain both the lack of folding in fragments of AS-DHFR containing fewer than 86 amino acids and the folding of the larger 37–159 and, possibly, 1–107 fragments. In this model, the principal driving force for folding derives from the exclusion of water from nonpolar side chains that become buried in the interior of a folded protein (Kauzmann, 1959). The apparent requirement for a sequence length in excess of 86 amino acids for AS-DHFR fragments might reflect the requirement for a minimal hydrophobic core. For this core, the entropy gained from the release of water from buried hydrophobic surfaces would be sufficient to compensate for the entropy lost by ordering the backbone and the side chains.

Related studies on barnase (Kippen et al., 1994) and chymotrypsin inhibitor-2 (de Prat Gay & Fersht, 1994; de Prat Gay et al., 1994) have shown that fragments of these two small proteins have little tendency to fold in aqueous solution. Particularly pertinent to the present study, Fersht and his colleagues (Itzhaki et al., 1995) constructed a series of peptides encompassing a putative α-helical nucleation site for CI-2 inhibitor. Similar to the present results, the successive inclusion of several adjacent elements of secondary structure did not stabilize the α-helix significantly. Long-range interactions with the remainder of the protein were postulated to be essential in stabilizing the nucleation site.

Relationship to kinetic intermediates in DHFR

Pulse-labeling hydrogen exchange NMR (Jones & Matthews, 1995) studies have shown previously that a large fraction of the earliest detectable folding intermediates for full-length DHFR has a native-like β-sheet topology that includes both structural domains. The integration of the sequences corresponding to both the ABD and the large domain into a global entity at this early stage of folding may reflect, in part, the lack of stability in the isolated domains. If either of these domains had significant stability, it could act as a kinetic trap that might retard the development of the β-sheet (Sosnick et al., 1994).

The observation of cooperatively folded structure in the 37–159 fragment, even in the absence of Am_2SO_4 (Figs. 3, 5), raises the question of whether partially or fully folded forms of this fragment act as a trap in the folding of the full-length protein. As described above, NMR results show that strand β1, contained in the 1–36 segment, is already inserted into the middle of the sheet within 13 ms for approximately half of the population of the early folding intermediate (Jones & Matthews, 1995). The remaining half does not protect its amide hydrogens against exchange in this time frame, implying either that the β-sheet has not formed, or that it forms a nonnative structure that must rearrange in subsequent folding reactions. Conceivably, this misfolded form could include the structure adopted by 37–159.

A possible mechanistic explanation for the two populations of early folding intermediates detected by NMR is that there are two nucleation sites in DHFR. The helix, loop, and strand contained in the 1–36 segment form a compact unit that may act as a nucleation site (Fersht, 1995; Itzhaki et al., 1995) for the formation of the native-like sheet. The limited solubility of the 1–36 fragment (C.V. Gegg & C.R. Matthews, unpubl. results) suggests that a significant driving force for its early association with the remainder of the protein is the hydrophobic effect (Kauzmann, 1959). The interaction of this segment with the remainder of the protein, however, could be in competition with another potential nucleation site in the ABD, residues 37–87, that directs folding to an alternative, stable conformation. The distortions in the sheet required to convert the nonnative conformer to the native conformer by inserting strand β1 into the interior of the sheet might be sufficient to allow exchange of the amide hydrogens.

Overview on fragmentation studies

The observation of a structurally and thermodynamically well-defined fold for fragment 37–159 from AS-DHFR is a surprising result that may provide an explanation for the previously observed heterogeneity in early folding intermediates of the full-length protein (Jones & Matthews, 1995). The potential opportunity to create and examine nonnative folds (Chaffotte et al., 1991; Hamada et al., 1996) provides a unique advantage for the fragmentation approach that complements the power of the native-state hydrogen exchange experiment (Bai et al., 1995) used to study native-like folding intermediates in full-length proteins.

Materials and methods

Reagents

NTCB was obtained from Sigma (St. Louis, Missouri) and Chelating Sepharose Fast Flow resin was purchased from Pharmacia (Piscataway, New Jersey). All other reagents were reagent grade or better.

Fragment preparation

A plasmid encoding a cysteine-free form of DHFR, C85A/C152S (Iwakura et al., 1995), designated AS DHFR, was the platform for cysteine mutagenesis. PCR-mediated mutagenesis (Sarker & Sommer, 1990), polyhistidine affinity tagging (Hochuli et al., 1988),

and cysteine-specific chemical cleavage with NTCB (Jacobson et al., 1973) were combined to enhance the production and isolation of protein fragments.

Briefly, three parent vectors encoding the cysteine-free AS DHFR with polyhistidine affinity tags (His 6) engineered at either or both termini were hosts for various cysteine point mutations (N37C, D87C, Q108C). The point mutations were engineered either singly into vectors tagged at either termini or, in the case of the doubly tagged vectors, in pairs (N37C/Q108C or N37C/D87C). The mutant, full-length fusion proteins were expressed in *E. coli*, purified by immobilized metal ion affinity chromatography on an Ni^+-IDA resin, and cleaved chemically at cysteine by NTCB. The untagged fragment was isolated subsequently with a second round of IMAC. Single cysteine mutants in combination with the appropriate single affinity tag allow the isolation of either the amino or carboxy fragments, whereas the double cysteine mutants with double affinity tags provide access to internal fragments. Further details on the methods utilized for fragment preparation are described in Gegg et al. (1996).

Physiochemical characterization

Fragment purity was assessed by SDS-PAGE (Schagger & von Jagow, 1987) and reverse-phase HPLC (Gegg et al., 1996) and determined to be >95%. The molecular mass of each fragment was measured by electrospray mass spectrometry and found to be within ±0.05% of the calculated value (Gegg et al., 1996).

Optical spectroscopy

All spectroscopic data were collected at 15 °C. The CD and fluorescence spectra were collected from samples dissolved in 10 mM potassium phosphate, pH 7.8, 0.2 mM potassium EDTA, and 1 mM β-mercaptoethanol. The extinction coefficients of the fragments were estimated by the method of Gill and von Hippel (1989) and were calculated to be: 1–36, 11,380 M^{-1} cm^{-1}; 1–86, 22,760 M^{-1} cm^{-1}; 1–107, 24,040 M^{-1} cm^{-1}; 37–159, 22,190 M^{-1} cm^{-1}; 87–159, 10,810 M^{-1} cm^{-1}; 108–159, 9,530 M^{-1} cm^{-1}; 37–86, 11,380 M^{-1} cm^{-1}; and 37–107, 12,660 M^{-1} cm^{-1}. The estimated error in these values is ±10%.

CD spectra were collected on an Aviv spectrometer, model 62DS. Far-UV CD spectra were scanned from 320 to 185 nm using 1-mm or 2-mm cuvettes with DHFR fragment concentrations of 4.1–18.8 μM. Near-UV CD spectra of 37–159 were scanned from 320 to 240 nm using a 10-cm cuvette and a fragment concentration of 7.7 μM. Raw data were converted to mean residue ellipticity by

$$MRE = (\Theta * 100 * MW)/(C * D * N_A),$$

where Θ is the ellipticity in degrees, MW is the molecular weight of the given fragment, C is the concentration in mg mL^{-1}, D is the pathlength in cm, and N_A is the number of residues in the fragment.

Fluorescence spectra were collected with a Spex Fluorolog spectrometer, model 1681, or an Aviv model ATF105 spectrofluorometer using a 1-cm cuvette. The excitation wavelength was fixed at 295 nm and the emission spectra scanned from 305 to 460 nm. The CMW was calculated for each spectrum using the equation

$$CMW = \sum \lambda_i * I_i / \sum I_i$$

where λ_i is wavelength i and I_i is the fluorescence intensity at λ_i. CMW is chosen to monitor the unfolding transition, as opposed to the maximum emission wavelength, to take into account the changes in spectral shape that occur as a result of denaturation. Fluorescence spectral asymmetries result in CMW values that are shifted compared with the maximum emission wavelength.

Chemical denaturation experiments

Fragments were dialyzed against 10 mM potassium phosphate, pH 7.8, 0.2 mM potassium EDTA, and 1 mM β-mercaptoethanol. Fragments were then diluted to varying final urea concentrations in the above buffer and equilibrated for 30 min at 15 °C prior to data collection. The signal was found to be invariant after 30 min, demonstrating that equilibrium had been reached. Final fragment concentrations ranged from 1.9 to 9.5 μM. The pathlength for fluorescence was 1 cm; for far-UV CD, 0.2 or 0.5 cm; and for near-UV CD, 10 cm.

Urea-induced unfolding curves of fragment 37–159 and the full-length AS DHFR protein were analyzed by calculating the fraction apparent unfolded, F_{app}, as a function of urea using the equation

$$F_{app} = (Y_f - Y_o)/(Y_f - Y_u),$$

where Y_o is the observed signal, either CD (222 nm or 292 nm) or fluorescence (CMW), at any given urea concentration, and Y_f and Y_u are the values of the folded and unfolded baselines, respectively, at the same urea concentration (Pace, 1986; Pace et al., 1990; Finn et al., 1991).

Thermodynamic parameters were extracted from these data by analyzing F_{app} curves with the Savuka software package version 5.0 (D. Lambright & O. Bilsel, pers. comm.; Zitzewitz et al., 1995). Assuming that the free energy of folding varies linearly with denaturant concentrations (Schellman, 1978), the free energy change in the absence of denaturant, $\Delta G°(H_2O)$, the transition midpoint, C_m, and the m value, a measure of the sensitivity of the unfolding transition to denaturant, were calculated for a two-state folding model as described previously (Finn et al., 1991). In the global analysis of multiple data sets, the $\Delta G°(H_2O)$ and the m value were linked parameters.

Acknowledgments

This research was supported by NSF grant MCB9604678 to C.R.M. and NIH grant GM16516 to C.V.G. We are grateful to Drs. Ronald Wetzel and Jill Ann Zitzewitz for helpful discussions. We also thank Osman Bilsel for his development of the Savuka data analysis method.

References

Alexandrescu AT, Abeygunawardana C, Shortle D. 1994. Structure and dynamics of a denatured 131-residue fragment of staphylococcal nuclease: A heteronuclear NMR study. *Biochemistry 33*:1063–1072.

Bai Y, Englander SW. 1996. Future directions in folding: The multi-state nature of protein structure. *Proteins Struct Funct Genet 24*:145–151.

Bai Y, Sosnick TR, Mayne L, Englander SW. 1995. Protein folding intermediates: Native-state hydrogen exchange. *Science 269*:192–197.

Baldwin RL. 1991. Molten globules: Specific or nonspecific folding intermediates? *Chemtracts—Biochemistry and molecular biology 2*:379–389.

Bystroff C, Oatley SJ, Kraut J. 1990. Crystal structures of *Escherichia coli* dihydrofolate reductase: The NADP+ holoenzyme and the folate–NADP+ ternary complex. Substrate binding and a model for the transition state. *Biochemistry 29*:3263–3277.

Chaffotte A, Guillou Y, Delepierre M, Hinz HJ, Goldberg ME. 1991. The isolated C-terminal (F2) fragment of the *Escherichia coli* tryptophan synthase beta-2 subunit folds into a stable, organized nonnative conformation. *Biochemistry 30*:8067–8074.

Chamberlain AK, Handel TM, Marqusee S. 1996. Detection of rare partially folded molecules in equilibrium with the native conformation of RNaseH. *Nature Struct Biol 9*:782–787.

Dabora JM, Pelton JG, Marqusee S. 1996. Structure of the acid state of *Escherichia coli* ribonuclease HI. *Biochemistry 35*:11951–11958.

de Prat Gay G, Fersht AR. 1994. Generation of a family of protein fragments for structure-folding studies. 1. Folding complementation of two fragments of chymotrypsin inhibitor-2 formed by cleavage at its unique methionine residue. *Biochemistry 33*:7957–7963.

de Prat Gay G, Ruiz-Sanz J, Fersht AR. 1994. Generation of a family of protein fragments for structure-folding studies. 2. Kinetics of association of the two chymotrypsin inhibitor-2 fragments. *Biochemistry 33*:7964–7970.

Dill KA. 1985. Theory for the folding and stability of globular proteins. *Biochemistry 24*:1501–1509.

Dill KA, Alonso DOV, Hutchinson K. 1989. Thermal stabilities of globular proteins. *Biochemistry 28*:5439–5449.

Dobson CM. 1994. Protein folding: Solid evidence for molten globules. *Curr Biol 4*:636–640.

Dyson HJ, Merutka G, Waltho JP, Lerner RA, Wright PE. 1992a. Folding of peptide fragments comprising the complete sequence of proteins. Models for initiation of protein folding. I. Myohemerythrin. *J Mol Biol 226*:795–817.

Dyson HJ, Sayre JR, Merutka G, Shin HC, Lerner RA, Wright PE. 1992b. Folding of peptide fragments comprising the complete sequence of proteins. Models for initiation of protein folding. II. Plastocyanin. *J Mol Biol 226*:819–835.

Eliezer D, Wright PE. 1996. Is apomyoglobin a molten globule? Structural characterization by NMR. *J Mol Biol 263*:531–538.

Falzone CJ, Mayer MR, Whiteman EL, Moore CD, Lecomte JTJ. 1996. Design challenges for hemoproteins: The solution structure of apocytochrome b5. *Biochemistry 35*:6519–6526.

Fersht AR. 1995. Optimization of rates of protein folding: The nucleation-condensation mechanism and its implications. *Proc Natl Acad Sci USA 92*:10869–10873.

Finn BE, Chen X, Jennings PA, Saalau-Bethell SM, Matthews CR. 1991. Principles of protein stability. Part 1—Reversible unfolding of proteins: Kinetic and thermodynamic analysis. In: Rees AR, Wetzel R, Sternberg JE, eds. *Protein engineering—A practical approach*. Oxford: IRL Press. 167–189.

Gegg CV, Bowers KE, Matthews CR. 1996. A general approach for the design and isolation of protein fragments: The molecular dissection of dihydrofolate reductase. In: Marshak D, ed. *Techniques in protein chemistry VII*. San Diego, California: Academic Press. pp 439–448.

Gill SC, von Hippel PH. 1989. Calculation of protein extinction coefficients from amino acid sequence data. *Anal Biochem 182*:319–326.

Hamada D, Segawa S, Goto Y. 1996. Non-native alpha helical intermediate in the refolding of beta-lactoglobulin, a predominantly beta-sheet protein. *Nature Struct Biol 3*:868–873.

Hochuli E, Bannwarth W, Dobeli H, Gentz R, Stuber D. 1988. Genetic approach to facilitate purification of recombinant proteins with a novel metal chelate adsorbent. *Bio/Technology 6*:1321–1325.

Itzhaki LS, Neira JL, Ruiz-Sanz J, de Prat Gay G, Fersht, AR. 1995. Search for nucleation sites in smaller fragments of chymotrypsin inhibitor-2. *J Mol Biol 254*:286–304.

Iwakura M, Jones BE, Luo J, Matthews CR. 1995. A strategy for testing the suitability of cysteine replacements in dihydrofolate reductase from *Escherichia coli*. *J Biochem 117*:480–488.

Jacobson GR, Schaffer MH, Stark GR, Vanaman TC. 1973. Specific chemical cleavage in high yield at the amino peptide bonds of cysteine and cystine residues. *J Biol Chem 248*:6583–6591.

Jaenicke R. 1991. Protein folding: Local structures, domains, subunits, and assemblies. *Biochemistry 30*:3147–3161.

Jones BE, Matthews CR. 1995. Early intermediate in the folding of dihydrofolate reductase from *Escherichia coli* detected by hydrogen exchange and NMR. *Protein Sci 4*:167–177.

Kauzmann W. 1959. Some factors in the interpretation of protein denaturation. *Adv Protein Chem 14*:1–63.

Kim PS, Baldwin RL. 1982. Specific intermediates in the folding reactions of small proteins and the mechanism of protein folding. *Annu Rev Biochem 51*:459–489.

Kippen AD, Sancho J, Fersht AR. 1994. Folding of barnase in parts. *Biochemistry 33*:3778–3786.

Kuwajima K. 1996. The molten globule state of alpha-lactalbumin. *FASEB J 10*:102–109.

Kuwajima K, Garvey EP, Finn BE, Matthews CR, Sugai S. 1991. Transient intermediates in the folding of dihydrofolate reductase as detected by far-ultraviolet circular dichroism spectroscopy. *Biochemistry 30*:7693–7703.

Lecomte JTJ, Kao YH, Cocco MJ. 1996. The native state of apomyoglobin described by proton NMR spectroscopy: The A-B-G-H interface of wild-type sperm whale apomyoglobin. *Proteins Struct Funct Genet 25*:267–285.

Mann CJ, Royer CA, Matthews CR. 1993. Tryptophan replacement in the trp aporepressor from *Escherichia coli*: Probing the equilibrium and kinetic folding models. *Protein Sci 2*:1853–1861.

Matthews CR. 1993. Pathways of protein folding. *Annu Rev Biochem 62*:653–683.

Missiakas D, Betton JM, Minard P, Yon JM. 1990. Unfolding-refolding of the domains in yeast phosphoglycerate kinase: Comparison with the isolated engineered domains. *Biochemistry 29*:8683–8689.

Oas TG, Kim PS. 1988. A peptide model of a protein folding intermediate. *Nature 336*:42–48.

Pace CN. 1986. Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods Enzymol 131*:266–280.

Pace CN, Shirley BA, Thomson JA. 1990. Measuring the conformational stability of a protein. In: Creighton TE, ed. *Protein structure—A practical approach*. Oxford: IRL Press. 311–330.

Peng Z, Kim PS. 1994. A protein dissection study of a molten globule. *Biochemistry 33*:2136–2141.

Raschke TM, Marqusee S. 1997. The kinetic folding intermediate of ribonuclease H resembles the acid molten globule and partially unfolded molecules detected under native conditions. *Nature Struct Biol 4*:298–304.

Richardson JS. 1981. The anatomy and taxonomy of protein structure. *Adv Protein Chem 34*:168–339.

Sarker G, Sommer S. 1990. The "Megaprimer" method of site-directed mutagenesis. *BioTechniques 8*:404–407.

Sawaya MR, Kraut J. 1997. Loop and subdomain movements in the mechanism of *Escherichia coli* dihydrofolate reductase: Crystallographic evidence. *Biochemistry 36*:586–603.

Schagger H, von Jagow G. 1987. Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Anal Biochem 166*:368–379.

Schellman JA. 1978. Solvent denaturation. *Biopolymers 17*:1305–1322.

Shortle D, Meeker AK. 1986. Mutant forms of staphylococcal nuclease with altered patterns of guanidine hydrochloride and urea denaturation. *Proteins Struct Funct Genet 1*:81–89.

Shortle D, Wang Y, Gillespie JR, Wrabl JO. 1996. Protein folding for realists: A timeless phenomenon. *Protein Sci 5*:991–1000.

Skolnick J, Kolinski, A, Godzik A. 1993. From independent modules to molten globules: Observations on the nature of protein folding intermediates. *Proc Natl Acad Sci USA 90*:2099–2100.

Sosnick TR, Mayne L, Hiller R, Englander SW. 1994. The barriers in protein folding. *Nature Struct Biol 1*:149–156.

Staley JP, Kim PS. 1994. Formation of a native-like subdomain in a partially folded intermediate of bovine pancreatic trypsin inhibitor. *Protein Sci 3*:1822–1832.

Tasayco ML, Carey J. 1992. Ordered self-assembly of polypeptide fragments to form nativelike dimeric trp repressor. *Science 255*:594–597.

Wu LC, Grandori R, Carey J. 1994. Autonomous subdomains in protein folding. *Protein Sci 3*:369–371.

Wu LC, Laub PB, Elove GA, Carey J, Roder H. 1993. A noncovalent peptide complex as a model for an early folding intermediate of cytochrome c. *Biochemistry 32*:10271–10276.

Zitzewitz J, Bilsel O, Luo J, Jones BE, Matthews CR. 1995. Probing the folding of a leucine zipper peptide by stopped-flow circular dichroism spectroscopy. *Biochemistry 34*:12812–12819.