Probing the folding pathway of a β-clam protein with single-tryptophan constructs

Patricia L Clark[1,2,3], Benjamin F Weston[1,4] and Lila M Gierasch[1]

Background: Cellular retinoic acid binding protein I (CRABPI) is a small, predominantly β-sheet protein with a simple architecture and no disulfides or cofactors. Folding of mutants containing only one of the three native tryptophans has been examined using stopped-flow fluorescence and circular dichroism at multiple wavelengths.

Results: Within 10 ms, the tryptophan fluorescence of all three mutants shows a blue shift, and stopped-flow circular dichroism shows significant secondary structure content. The local environment of Trp7, a completely buried residue located near the intersection of the N and C termini, develops on a 100 ms time scale. Spectral signatures of the other two tryptophan residues (87 and 109) become native-like in a 1 s kinetic phase.

Conclusions: Formation of the native β structure of CRABPI is initiated by rapid hydrophobic collapse, during which local segments of chain adopt significant secondary structure. Subsequently, transient yet specific interactions of amino acid residues restrict the arrangement of the chain topology and initiate long-range associations such as the docking of the N and C termini. The development of native tertiary environments, including the specific packing of the β-sheet sidechains, occurs in a final, highly cooperative step simultaneous with stable interstrand hydrogen bonding.

Addresses: [1]Department of Chemistry, University of Massachusetts, Amherst, MA 01003, USA. [2]Molecular Biophysics Graduate Program, University of Texas Southwestern Medical Center, Dallas, TX 75235, USA.

Present addresses: [3]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [4]University of Utah School of Medicine, Salt Lake City, UT 84132, USA.

Correspondence: Lila M Gierasch
E-mail: gierasch@chem.umass.edu

Key words: β sheet, fluorescence, folding, mutations, tryptophan

Received: 28 May 1998
Revisions requested: 29 June 1998
Revisions received: 31 July 1998
Accepted: 24 August 1998

Published: 30 September 1998
http://biomednet.com/elecref/1359027800300401

Folding & Design 30 September 1998, 3:401–412

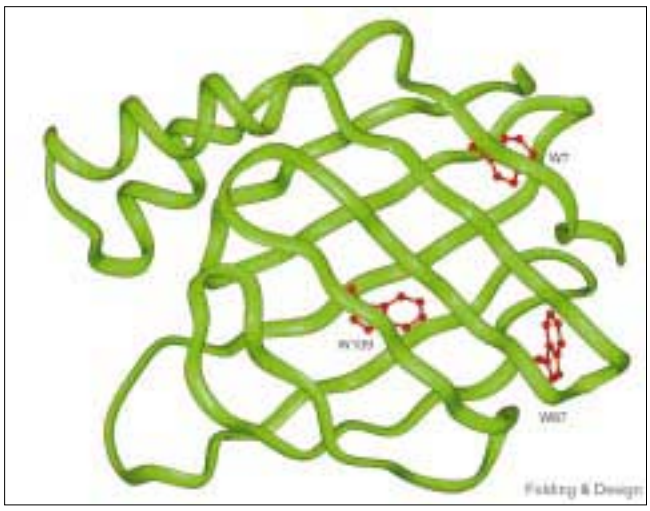
Introduction

Analyses of the folding of a wide range of helix-rich proteins, closely related mutant versions (for examples, see [1–3]) and a multitude of peptide studies of the helix/coil transition have shed considerable light on the folding of α helices [4,5]. In contrast, relatively little is known about the structural basis for β-sheet formation [6]. Recent work has made great strides in predicting the β-sheet propensity of unknown protein sequences [7,8], and in designing *de novo* β-sheet structures [9]. Numerous questions remain concerning the fundamental mechanism for the folding of β sheets. Is sheet formation generally initiated via the same hydrophobic collapse often seen for helical proteins [10–14]? Would hydrophobic collapse provide the polypeptide chain with enough protection from intermolecular association, or are other structural intermediates (for example, α helices) populated before β-sheet formation [15,16]? Do specific clusters of residues have a role in nucleating folding [17,18]? What are the roles of specific tertiary interactions in determining the native structure? When do hydrogen bonds form, and how cooperative is the process?

CRABPI is a 136-residue cytoplasmic protein consisting of a 10-stranded β-clam structure and a small segment of helix–turn–helix (Figure 1); there is a large central cavity inside the barrel [19] that is solvent-filled in the apoprotein structure [20]. It is a member of the large family of intracellular lipid-binding proteins (iLBPs), all of which fold into strikingly similar native structures despite sequence homology ranging from 18 to 80% [21]. Consistent with this sequence variability, CRABPI is very tolerant to amino-acid substitution while retaining its native structure and binding ability [22,23]. CRABPI has no disulfide bonds, cofactors, or post-translational modifications, and the equilibrium folding behavior of apo-CRABPI is well characterized [24]. The amino acid sequence contains three tryptophan residues, at positions 7, 87 and 109.

CRABPI provides several distinct advantages as a protein folding model. Its simple single-sheet architecture facilitates interpretation of folding results. For example, our laboratory has demonstrated that chain topology develops almost a full order of magnitude faster ($\tau \sim 200$ ms) than the formation of the native hydrogen-bond network ($\tau = 1$ s) [25]. The three tryptophan residues in the sequence are at structurally distinct locations, thereby providing three regiospecific structural probes for fluorescence studies. The inherent tolerance of the structure to amino acid substitutions [22] permits complete deconvolution of the fluorescence signals during folding by the

Figure 1



Ribbon diagram of the crystal structure of apo-CRABPI, from Thompson *et al.* [20]. The locations of the three tryptophan residues are shown in red.

construction of variants containing a single Trp. Note the unique environments of these tryptophan residues (Figure 1). Trp7 is completely buried and is in close contact with the sidechains of several non-polar, non-aromatic residues and with the sidechain of Arg135; an aromatic residue at position 7 is conserved throughout the iLBP family, as is the specific contact between the N- and C-terminal strands mediated by Arg135 [21]. Trp87 is the most exposed of the three tryptophan residues, with 30% of its surface area in contact with the solvent. One side of the ring structure, however, is near to the sidechains of two phenylalanine residues. Trp87 is not as well conserved as Trp7, appearing only in CRABP (I and II) and cellular retinol-binding proteins (CRBPI and II). Trp109 is almost completely protected from solvent and points in towards the ligand-binding cavity; it interacts with three phenylalanine sidechains, and is close to the sulfur of Cys95, which lies over the indole ring. Conservation of Trp109 is identical to that for Trp87.

Our laboratory has previously reported the construction of 'kinetic folding spectra' derived from stopped-flow fluorescence (SF-Flu) data collected during the folding of wild-type CRABPI [26]. SF-Flu traces were acquired at seven discrete wavelengths, and these data were used to construct full spectra during the folding process. Each trace showed the existence of three observable phases (at 0.1 s, 1 s and ~20 s), plus at least one additional phase during the dead time of the stopped-flow mixing. Examination of the full emission spectrum during folding clearly demonstrated that the dead time changes place one or more tryptophan sidechains in a more hydrophobic environment than is experienced in the unfolded protein, and suggested that the dead-time folding process is accompanied by concomitant rearrangement of all three tryptophan environments.

Because of the presence of three tryptophan residues, the overlapping wild-type fluorescence results provided only some general characteristics of the folding intermediates. The relative contributions of each tryptophan residue to specific folding phases remained unclear, and it was not possible to determine with certainty which parts of the amino acid sequence were involved in each phase. An initial series of single-Trp substitutions [26] indicated the relative contributions of each tryptophan residue to the native fluorescence spectrum. We speculated that the early increase in fluorescence intensity arose from sequestration of Trp7, which has the greatest contribution to the low-wavelength side of the native fluorescence spectrum.

In the present study, we have further dissected the tryptophan spectral contributions by constructing stable, well folded CRABPI variants containing only one tryptophan residue. SF-Flu studies of these proteins have enabled complete deconvolution of the fluorescence of CRABPI during folding. From kinetic traces and spectra collected during the folding of these mutant proteins, each fluorescence phase observed during wild-type CRABPI folding was assigned to the evolution of a specific interaction (or set of interactions). We have also collected stopped-flow circular dichroism (SF-CD) traces at multiple wavelengths during folding in order to determine the time scales for the formation of secondary structure and a tertiary aromatic interaction. Together these results have provided a much clearer understanding of the structure present at each stage of the CRABPI folding process, and suggest a general mechanism for folding of antiparallel, up-and-down β-barrel structure, such as that in the iLBP family. This involves initial hydrophobic collapse, followed by formation of a solvated state with native-like topology but without stable hydrogen bonds, and then cooperative development of the stable hydrogen-bonding network and tertiary packing.

Results

Single-Trp mutant proteins have unaltered CRABPI structural characteristics

Expression and purification of mutant proteins

All single-Trp mutants were constructed in a pseudo-wild-type CRABPI background (WT*) that contained an Arg131→Gln mutation for increased stability [22] and an N-terminal His tag to facilitate purification (see Materials and methods for a complete description of CRABPI constructs). WT* overexpressed well in *Escherichia coli*, partitioning to the cytoplasm; purification was therefore a one-step process using a Ni-NTA column (see Materials and methods). A typical yield for purified WT* was 35 mg/l cell culture. Although expression levels for each Trp-mutant protein were similar to that of WT* (data not shown), the yield from

the refolding step varied. A high yield of refolded protein was used as a criterion of mutant protein stability when discriminating between two potential versions of a particular single-Trp CRABPI variant (Table 1); for example, preparation of proteins containing only Trp7 (Trp87 mutated to either Phe or Tyr, Trp109 mutated to Tyr) resulted in refolded solutions that were either almost transparent (W87F) or virtually opaque (W87Y). Such a dramatic difference in refolding characteristics was not expected for two constructs with slight differences in only one residue location that is 30% exposed in the crystal structure of CRABPI [19]; it suggests that the W87Y-containing protein is much more prone to aggregation, presumably because of changes in the structural nature of the folding intermediates, or of their lifetimes during the folding process, or slight alterations in the final native structure. For these reasons, the Trp7-only construct used for additional studies was CRABPI-W87F/W109Y/R131Q(His). Using the same method, the Trp109-only construct selected was CRABPI-W7F/W87Y/R131Q(His). Typical yields for purified single-Trp proteins were approximately 20 mg/l cell culture.

CD spectra correspond to β-sheet proteins

Far-UV CD spectra for WT* and the three single-Trp CRABPI variants all show a minimum at 218 nm and a maximum at approximately 195 nm (Figure 2), which is characteristic of proteins consisting largely of β-sheet structure [27]. In addition, the CD spectrum for WT* contains an inflection at 228 nm. This has previously been shown to arise from fine details in the tertiary structure surrounding Trp87 and Trp109 [26]. Not surprisingly, the spectra for all three of the new Trp-mutant proteins (each lacking either Trp87 or Trp109, or both) are missing this inflection. When the proteins are heated, the spectra shift to a minimum at 203 nm, indicating a completely random-coil conformation [27].

Ligand binding with WT-like affinity*

Retinoic acid quenches CRABPI tryptophan fluorescence, which forms the basis of a method for estimating binding affinities of mutant proteins [22,26,28]. The three single-Trp CRABPI variants all bind retinoic acid with affinities comparable to that of WT* (Figure 3); that is, the titration curve has a break at stoichiometric (1:1) binding, with additional ligand having a significant, but greatly reduced, quenching effect (probably due to nonspecific interactions between the proteins and the highly hydrophobic ligand). For all four of the mutant proteins assayed, affinities for retinoic acid binding are not expected to be comparable to that of the wild-type protein [22], as the mutant proteins described here all have one of the primary ligand-binding residues (Arg131) mutated to a glutamine. This modification has previously been shown to markedly reduce the ligand-binding affinity of the protein [22].

Folding/unfolding is cooperative and reversible

Changes in the fluorescence spectra were used to monitor folding and unfolding transitions of WT* and the three single-Trp mutant proteins in urea (Figure 4). By this measure, WT* stability is comparable to that of CRABPI-R131Q [22], which suggests that the His tag does not perturb the structure. Each of the mutant proteins unfolded cooperatively and refolded reversibly, and the equilibrium data are well described by a two-state folding model [29]. The exception to this observation was Trp109-only, which has an equilibrium intermediate at

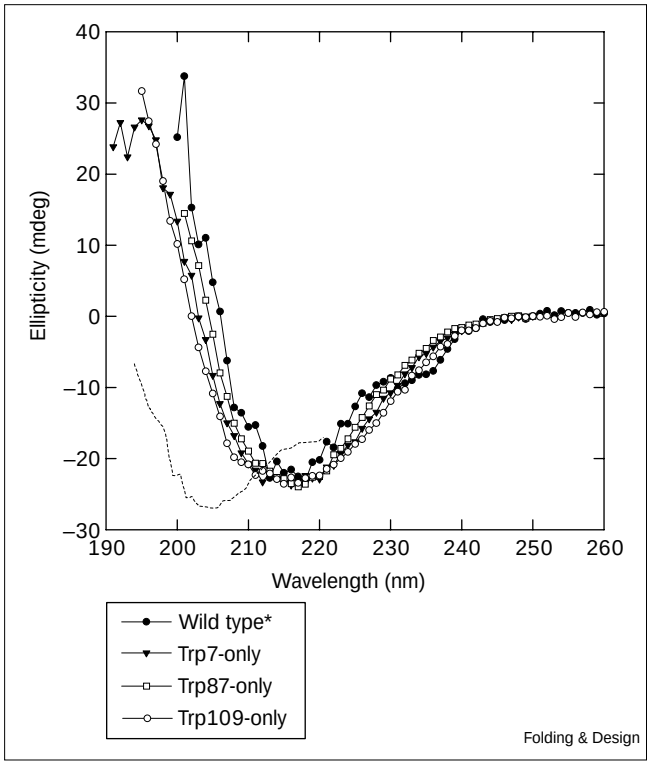
Table 1

Construction of CRABPI-WT* and single-Trp variants.

CRABPI construct name	Shortcut name†	Oligonucleotide used for construction‡	Template DNA
W109Y/R131Q(His)	–	CGTTGGCCAGCTCTCGGGTGTAGTA-AGTTTTAGGGCCATCCC	WT*
W87F/W109Y/R131Q(His)	Trp7-only	GCAGTGAATCTTGTTCTCATTCTC-GAACGTGGGTAAACTCCTGCATTTG	CRABPI-W109Y/R131Q(His)
W87Y/W109Y/R131Q(His)	–	GCAGTGAATCTTGTTCTCATTCTC-GTACGTGGGTAAACTCCTGCATTTG	CRABPI-W109Y/R131Q(His)
W7F/W109Y/R131Q(His)	Trp87-only	AATTCTCGCTGCTGCGCATCTTGAA-GGTACCGGCGAAGTTGGGC	CRABPI-W109Y/R131Q(His)
W7F/R131Q(His)	–	AATTCTCGCTGCTGCGCATCTTGAA-GGTACCGGCGAAGTTGGGC	WT*
W7F/W87F/R131Q(His)	–	GCAGTGAATCTTGTTCTCATTCTC-GAACGTGGGTAAACTCCTGCATTTG	CRABPI-W7F/R131Q(His)
W7F/W87Y/R131Q(His)	Trp109-only	GCAGTGAATCTTGTTCTCATTCTC-GTACGTGGGTAAACTCCTGCATTTG	CRABPI-W7F/R131Q(His)

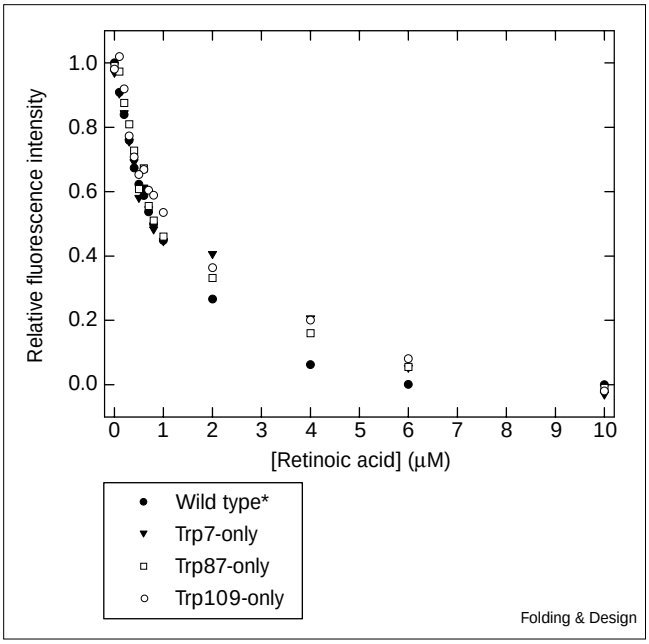
†Although all mutant proteins were expressed and purified, only those with shortcut names were subject to detailed analysis. ‡Complementary to coding strand; the Trp codon is bold. Another primer, complementary to the noncoding strand, was also used.

Figure 2



Far-UV CD spectra for WT* and single-Trp CRABPI variants. For each spectrum, protein concentration is 10 μM and the buffer is 10 mM sodium phosphate, pH 8.0, with 300 μM dithiothreitol. Spectra were collected at 5°C, and were baseline-corrected by subtracting a blank spectrum. A single spectrum is shown for denatured WT* (dotted line) at 80°C; spectra for denatured single-Trp proteins were equivalent.

Figure 3



Fluorimetric titration of WT* and single-Trp CRABPI variants with retinoic acid. Normalized fluorescence emission intensities are shown for each protein (1 μM) at 326 nm in 10 mM Tris pH 8.0 at room temperature, in the presence of increasing concentrations of retinoic acid.

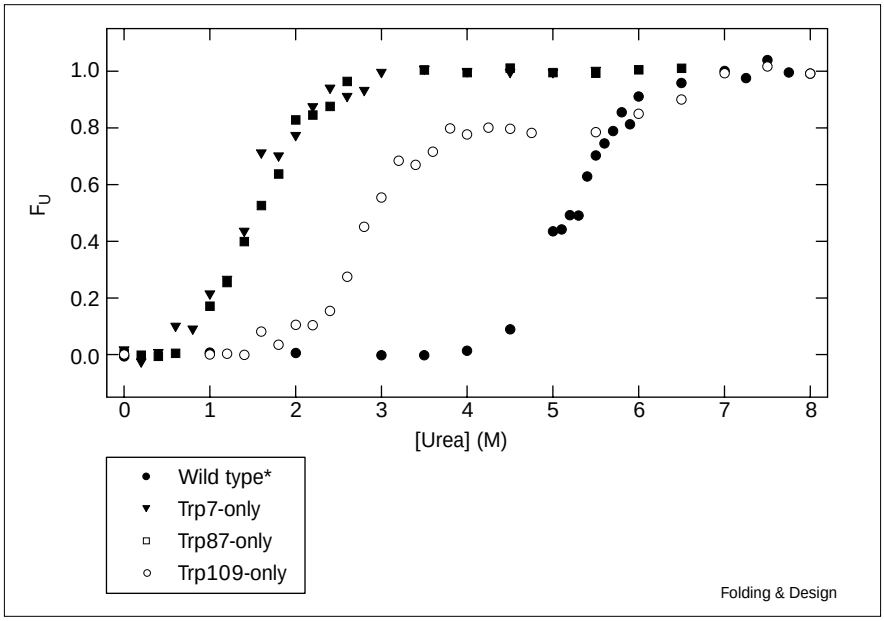
5.5 M urea. We are unsure at present as to the structural composition of this intermediate, but its fluorescence and CD spectra (data not shown) indicate that it does not contain an appreciable amount of either secondary or tertiary structure. In addition, under the kinetic conditions used in this study (< 2.4 M and > 5.5 M urea), this intermediate is never significantly populated. Whereas the single-Trp proteins are significantly less stable than WT*, all three are stable, completely folded proteins at urea concentrations up to 0.6 M. Furthermore, the reduced stability of the mutant proteins with respect to WT* is accounted for by substantially faster rates of unfolding with essentially unchanged folding kinetics (see below), enabling us to use them to dissect fluorescence changes upon CRABP folding.

Individual tryptophan fluorescence components

The fluorescence emission spectra for the unfolded proteins are similar, each with λ_{max} = 350 nm, indicating that the environments surrounding the tryptophan residues in each protein in 7 M urea are comparable to free tryptophan in solution (Figure 5a). The spectrum of folded WT* has a large reduction in intensity and a substantial (~20 nm) blue shift relative to the unfolded state, identical to the effect seen for wild-type CRABPI [26], where the intensity reduction was attributed to fluorescence quenching from residues surrounding Trp109 (Arg111 and Cys95) and the blue shift was explained by burial of the tryptophan sidechains. The three Trp-mutant proteins have unique fluorescence spectra. Trp7-only has most of its intensity centered at shorter wavelengths, and Trp87-only has more intensity at higher (~340 nm) wavelengths, but both proteins have fluorescence intensity that is enhanced in the native state as compared with the quenching seen for WT*. The third mutant protein, Trp109-only, has a native emission spectrum with λ_{max} shifted by approximately the same amount as Trp87-only, but there is substantially less intensity. The sum of the native fluorescence spectra for the three single-Trp proteins has a shape almost identical to that of WT* (Figure 5b). This result argues that the local environments around the tryptophan residues are unchanged relative to WT* and, therefore, that the mutant proteins have the same structure as WT*. The only significant difference between the spectrum of native WT* and the summed spectrum is that the latter has much greater overall intensity, indicating that there is energy transfer between the tryptophan residues in WT* [30].

Figure 4

Urea denaturation of WT* and single-Trp CRABPI variants. Denaturation was monitored as the change in the fluorescence emission intensity of each protein with increasing urea concentration, at the λ_{max} of each native protein. Raw fluorescence intensities were converted to the fraction of unfolded molecules (F_U) using the procedure described by Pace *et al*. [39].

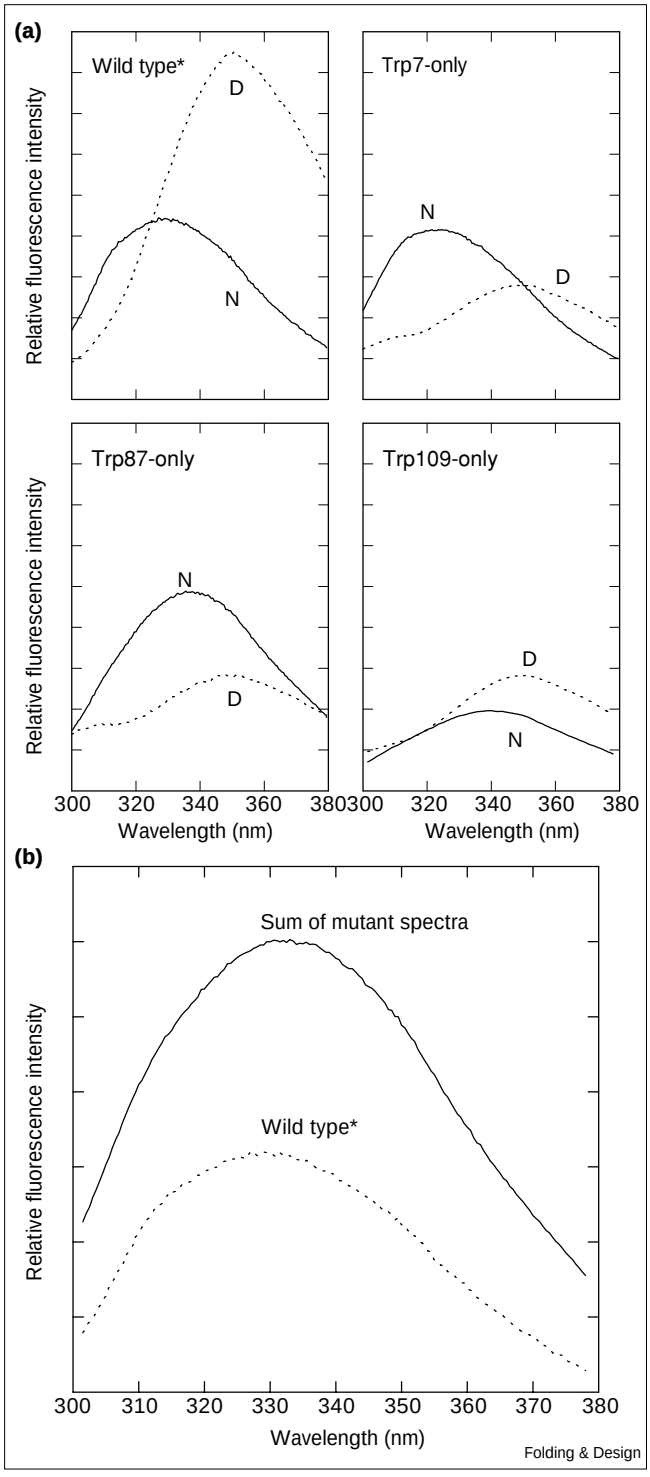


The fluorescence spectrum for each of the native single-Trp proteins (Figure 5a) also provides evidence of the environment surrounding each of the three tryptophan residues and correlates well with structural evidence gleaned from the crystal structures of CRABPI [19,20]. For example, the strong blue shift and high intensity for Trp7-only shows a tryptophan that is in a highly apolar and largely buried environment [31]. Trp87, in contrast, appears to have less solvent protection, as evidenced by its smaller blue shift, but still has a large intensity increase compared with the denatured spectrum (probably attributable to the two phenylalanine sidechains in contact with one face of the tryptophan sidechain). Trp109 has very little fluorescence in the native state, nor is the residual signal shifted much from the denatured λ_{max}, consistent with the close proximity of the tryptophan sidechain to the sidechains of Arg111 and Cys95. In addition, we have recently constructed a C95A mutant protein that clearly shows that the proximity of the Cys95 sidechain is the major factor in the quenching of Trp109 fluorescence (S. Eyles and L.M.G., unpublished results). These results were expected on the basis of our initial mutagenesis results [22,26] and reinforce the notion that these new mutant proteins, while now several steps removed from the original wild-type CRABPI, retain the overall native structure of CRABPI and, more importantly, retain the elements of the microenvironment surrounding each tryptophan residue.

Stopped-flow fluorescence at multiple emission wavelengths

SF-Flu of WT* folding shows kinetics that are virtually identical to wild-type CRABPI folding kinetics [26]; SF-Flu of the single-Trp proteins gives different results, depending on which tryptophan is present during folding (Figure 6). Kinetic traces were analyzed by fitting to a summed exponential equation, with the quality of fit determined by the randomness of residuals. Application of an autocorrelation function to the residuals demonstrated that a three-exponential fit had the highest randomness, indicating that the fluorescence change proceeds through three observable phases, with distinct time constants and amplitudes (Table 2). Also, there is an early dead-time change in the fluorescence spectrum that occurs before the first observable SF-Flu points. The folding of Trp7-only proceeds via a large dead-time change, which greatly increases the fluorescence intensity as well as shifting the λ_{max} by greater than 10 nm (Figure 6). The observable kinetic phases for Trp7-only are comparable to those for WT*; there are three phases, with a similar distribution of amplitudes (Table 2). Trp87-only, however, has a small dead-time fluorescence change, which has minor effects on the fluorescence intensity and results in a correspondingly small blue shift for λ_{max}. The fluorescence kinetics contain two observable phases, which correspond to the two slower phases seen for WT* and Trp7-only. Similarly, the observable phases of Trp109-only correspond well with these same two slow phases, but Trp109-only has a significant dead-time fluorescence change, more comparable to that seen for WT* and Trp7-only. The fluorescence intensity increases at short wavelengths and λ_{max} drops by > 10 nm. Trp7 is the only tryptophan that undergoes significant environmental changes during all four (dead-time plus three observable) kinetic phases. Trp87 has a minor dead-time change, and is largely uninvolved until late

Figure 5



Fluorescence emission spectra for WT* and single-Trp CRABPI variants. (a) Spectra under native (N: 10 mM Tris pH 8.0, solid line) and denaturing (D: 8 M urea, 10 mM Tris pH 8.0, dotted line) conditions. All spectra were collected at room temperature, with excitation at 280 nm. Protein concentrations were 6.7 μM. (b) Comparison of the native fluorescence emission spectrum of WT* with a spectrum summed from the spectra of the three single-Trp proteins.

(~1 s) in the progression of folding. Trp109, however, has an early change in its fluorescence signature but then does not encounter its final environment until the later phases of folding.

Whereas the amplitudes of the kinetic phases vary substantially for the mutant versions of CRABPI, there are no significant differences between the time constants for folding of the four proteins (Table 2). Conversely, there are large differences between the unfolding times (data not shown). For example, compared with WT* unfolding (370 s at 7 M urea), the unfolding time scale for the single-Trp CRABPI variants (1–7 s) is two orders of magnitude faster. Hence, the decreased stability of the three single-Trp proteins is reflected almost entirely in changes in the unfolding rates, enabling direct comparison of the folding rates for the four proteins.

Stopped-flow CD at multiple wavelengths

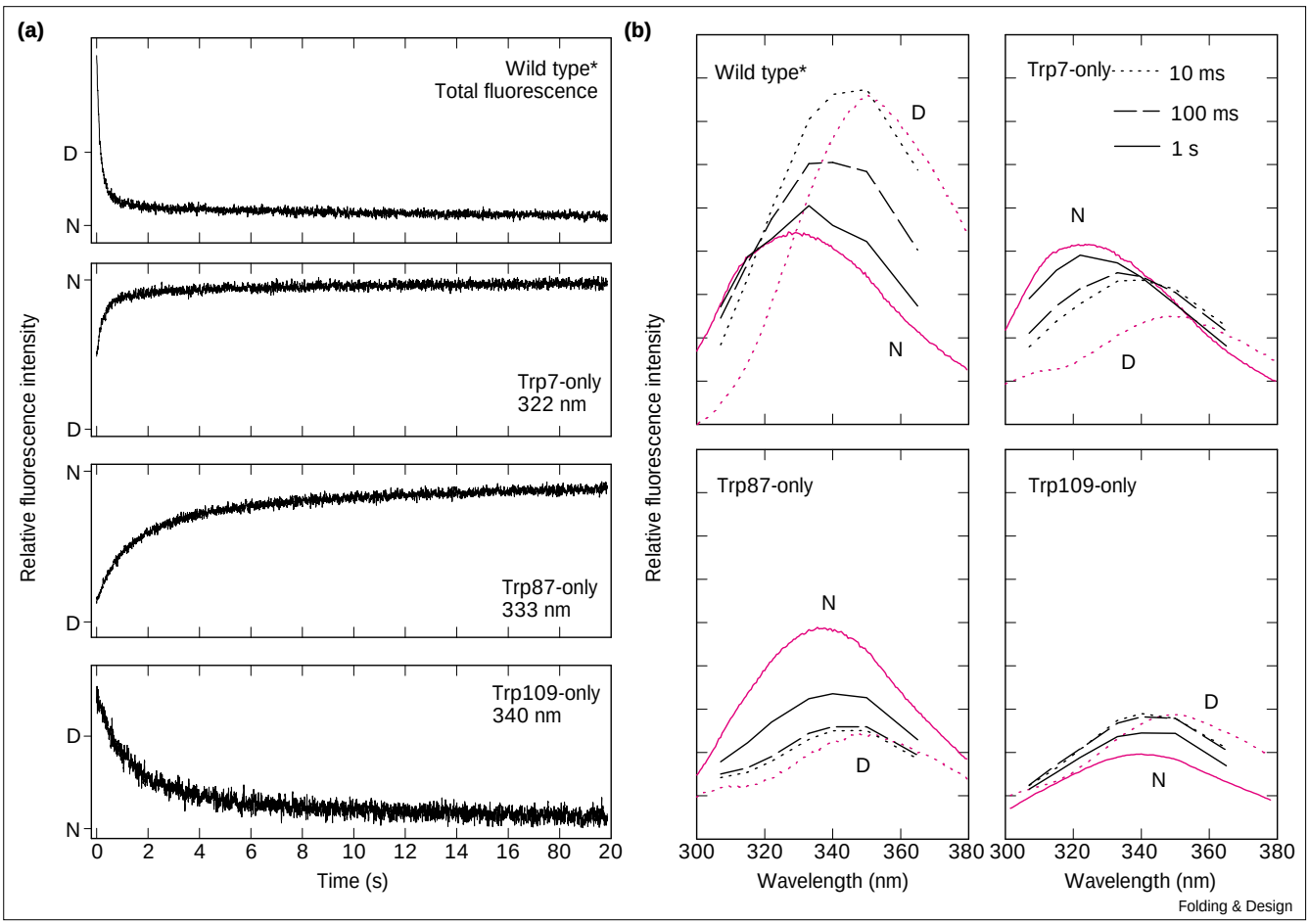
The dead-time spectrum

The SF-CD spectrum acquired after the dead time of WT* folding has substantial CD intensity, yet its shape is distinct from that of native CRABPI (Figure 7). Moreover, as folding proceeds to the native structure there continue to be significant changes in CD at multiple wavelengths (see below), arguing that the dead-time spectrum is non-native. The dominant feature in the dead-time spectrum is a minimum at 215 nm, shifted relative to the 218 nm minimum seen for the native spectrum. The 218 nm minimum is diagnostic for β structure, and its movement raises the possibility that the dead-time intermediate contains other secondary structure, such as increased amounts of random coil and/or possibly some α helix [15]. The dead-time spectrum contains a positive CD inflection centered at 224 nm. This is in contrast to the 228 nm inflection seen for native CRABPI, which arises from interactions contributing to the native environments surrounding Trp87 and Trp109 [26]. The inflection of the intermediate at 224 nm may therefore represent either modest modifications to the native environments or, alternatively, the presence of a totally non-native set of tertiary interactions. In addition, at low wavelengths (205–210 nm) the intermediate spectrum does not contain the positive CD contributions seen for the native spectrum. This lack of positive CD signal probably reflects an increased percentage of 'random coil' in the intermediate species.

Kinetics of WT folding in the observable time range*

Kinetic traces at two wavelengths — 205 nm, where ellipticity is positive in native CRABPI, consistent with its secondary structure content, and 232 nm, the location of the 'shoulder' in the native spectrum — are shown in Figure 8. At each wavelength, attaining native ellipticity involved only small changes from the ellipticity of the dead-time intermediate (Figure 7). Fitting the folding kinetics to a summed exponential equation is not possible

Figure 6



Kinetics of folding for WT* and single-Trp CRABPI variants. **(a)** Total fluorescence (WT*) or single-wavelength traces (single-Trp proteins) at the λ_{max} of each native protein. The relative intensities for native (N) and denatured (D) protein are indicated. **(b)** Constructed kinetic spectra, in red, with the native and denatured emission spectra described in Figure 5b (in black). Spectra were constructed at 10 ms (dotted red line), 100 ms (dashed red line) or 1 s (solid red line) after the initiation of folding.

given the low signal-to-noise levels. Qualitatively, however, the lower-wavelength kinetics (205 nm) appear to be dominated by a 100 ms phase, while higher-wavelength kinetics (232 nm) suggest near-equal contributions from 100 ms and ~1 s phases (Figure 8).

Discussion

The SF-Flu and SF-CD results described here have refined our understanding of the pathway by which CRABPI folds to its native state structure and have provided several clues as to the nature of the intermediates populated during folding. Our current model (Figure 9) incorporates the kinetic findings detailed above.

Folding is initiated via a global hydrophobic collapse

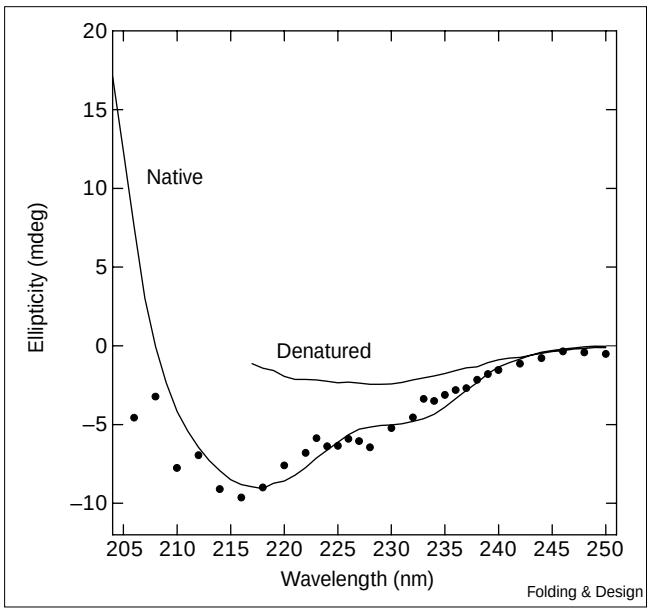
Evidence for a global hydrophobic collapse comes from the dead-time fluorescence changes seen for all three tryptophan residues. Even Trp87, which does not experience much dead-time intensity change, has an observable

Table 2

Summary of stopped-flow kinetic constants for the folding of wild-type and mutant versions of CRABPI monitored using fluorescence.

Experiment	Dead-time phase	Fast phase τ (ms)	Fast phase ΔA (%)	Medium phase τ (ms)	Medium phase ΔA (%)	Slow phase τ (ms)	Slow phase ΔA (%)
Wild type	++	100	70	1.2	15	20	15
WT*	++	200	80	1.1	10	18	10
Trp7-only	++	200	70	1.2	15	15	15
Trp87-only	+	–	–	1.0	60	16	40
Trp109-only	++	–	–	1.1	70	13	30

Figure 7



CD spectrum of the dead-time (approximately 10 ms) intermediate of CRABPI folding (filled circles). The error for each point is equal to twice the size of the circle. Native and denatured CRABPI CD spectra are shown for comparison. It was not possible to collect the denatured spectrum below 216 nm because of the large absorbance of urea at lower wavelengths. All three spectra have been subtracted from appropriate blank spectra.

blue shift in λ_{max}, indicating that it (as well as the other two Trp residues, which have larger λ_{max} shifts) is rapidly sequestered in a protected, hydrophobic environment. This early collapse increases Trp109 fluorescence to levels greater than those seen for native CRABPI, suggesting that burial of this tryptophan occurs before formation of interactions that lead to fluorescence quenching.

From the CD spectrum, it is clear that substantial secondary structure develops within the dead time, yet the locations of the CD bands suggest that the intermediate structure differs from that seen for native CRABPI, possibly by containing non-native helical elements but more probably due to additional random-coil contributions to the spectrum. This agrees well with evidence that the helix–turn–helix forms first during hydrogen/deuterium exchange/folding competition experiments (Z.P. Liu, J. Rizo and L.M.G., unpublished results), and the report that a peptide corresponding to the helix–turn–helix adopts to a significant extent a native-like conformation in aqueous solution [32]. The presence of a long-wavelength (224 nm) inflection indicates that the tryptophan environments are specific in the intermediate, and potentially packed with other aromatic sidechains, as this is the most typical cause of such CD bands [27]. Yet the displacement of the inflection relative to the native state (228 nm) demonstrates that the tryptophan environments do not yet contain all of their native components. A picture emerges of an early ensemble of structures, hydrophobically collapsed with the tryptophan sidechains sequestered from solvent and packed against other aromatic residues, and with a large amount of secondary structure, yet lacking the specific tertiary interactions that define the native state of CRABPI.

Folding proceeds through a step that alters the environment surrounding Trp7

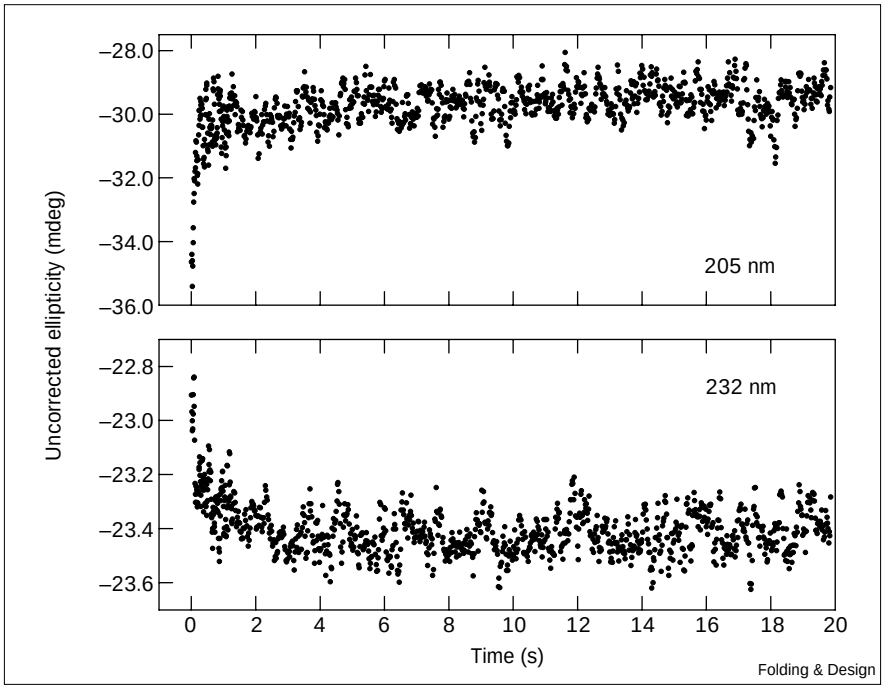
The next step in folding (τ = 200 ms for WT* family; 100 ms for WT) comprises the bulk (70%) of the observed fluorescence change for Trp7, and therefore is likely to be the point at which the sidechains of specific residues surrounding Trp7 in the native structure, including Arg135, approach this tryptophan sidechain. Hence this phase probably consists of the docking of the N- and C-terminal β strands. Interestingly, neither Trp87 nor Trp109 undergoes fluorescence-observable changes on this time scale; while this does not mean that there are no structural rearrangements in the neighborhood of these residues, it does indicate that any such adjustments do not have an effect on the overall nature of the environment surrounding Trp87 and Trp109. These results are interesting in light of the SF-CD kinetics. At the shorter wavelengths that reflect the β-sheet nature of the protein, it appears that the native secondary structure of CRABPI is developing on a 100–200 ms time scale. It should be emphasized that these structural arrangements are not evolving from denatured CRABPI, but rather from a dead-time population which already contains a substantial amount of secondary structure and a sub-native arrangement of the tryptophan environments. Hence the 100–200 ms phase may not reflect *de novo* β-sheet formation *per se*, but rather the arrangement of segments of preformed disorganized secondary structural elements into native β structure with proper topology. This is consistent with our previous finding that the central ligand-binding cavity of CRABPI, which is formed by the arrangement of the β strands, develops with a time constant of 200 ms, almost a full order of magnitude faster than the stable hydrogen-bond network of the β clamshell [25].

The environments of all three Trp residues change as they develop native tertiary contacts

In the next phase of folding (1 s time constant), the environments of all three Trp residues change to different degrees. At each position, the λ_{max} shifts to shorter wavelengths and the fluorescence spectrum gains more native-like intensity. Trp7 is already in a locale much like that in the native state, and presumably is experiencing minor adjustments of the positions of neighboring residues. Conversely, this folding phase represents the major portion of the observed fluorescence change for both Trp87 and Trp109 (60 and 70%, respectively). As a result, it is likely

Figure 8

Kinetics of CRABPI folding, monitored by SF-CD at two wavelengths, 205 and 232 nm. The total amplitude for each trace represents only a small fraction of the amplitude between denatured and native CRABPI.



to be the step at which Trp87 makes contact with the aromatic sidechains that are native-state neighbors, and Trp109 encounters its distinct native environment, including proximity to Cys95 in the neighboring strand.

Consistent with this explanation, we have previously demonstrated that the native protection from hydrogen exchange (indicating the formation of a stable hydrogen-bonding network) for amide protons in the β strands develops cooperatively on a 1 s time scale [25]. In addition, SF-CD experiments at longer wavelengths show an increased amplitude for the 1 s phase. The longer wavelengths often represent bands arising from substantial aromatic tertiary interactions, reinforcing the conclusion that the structural changes that result in the development of native tertiary interactions surrounding the tryptophan residues occur on a 1 s time scale.

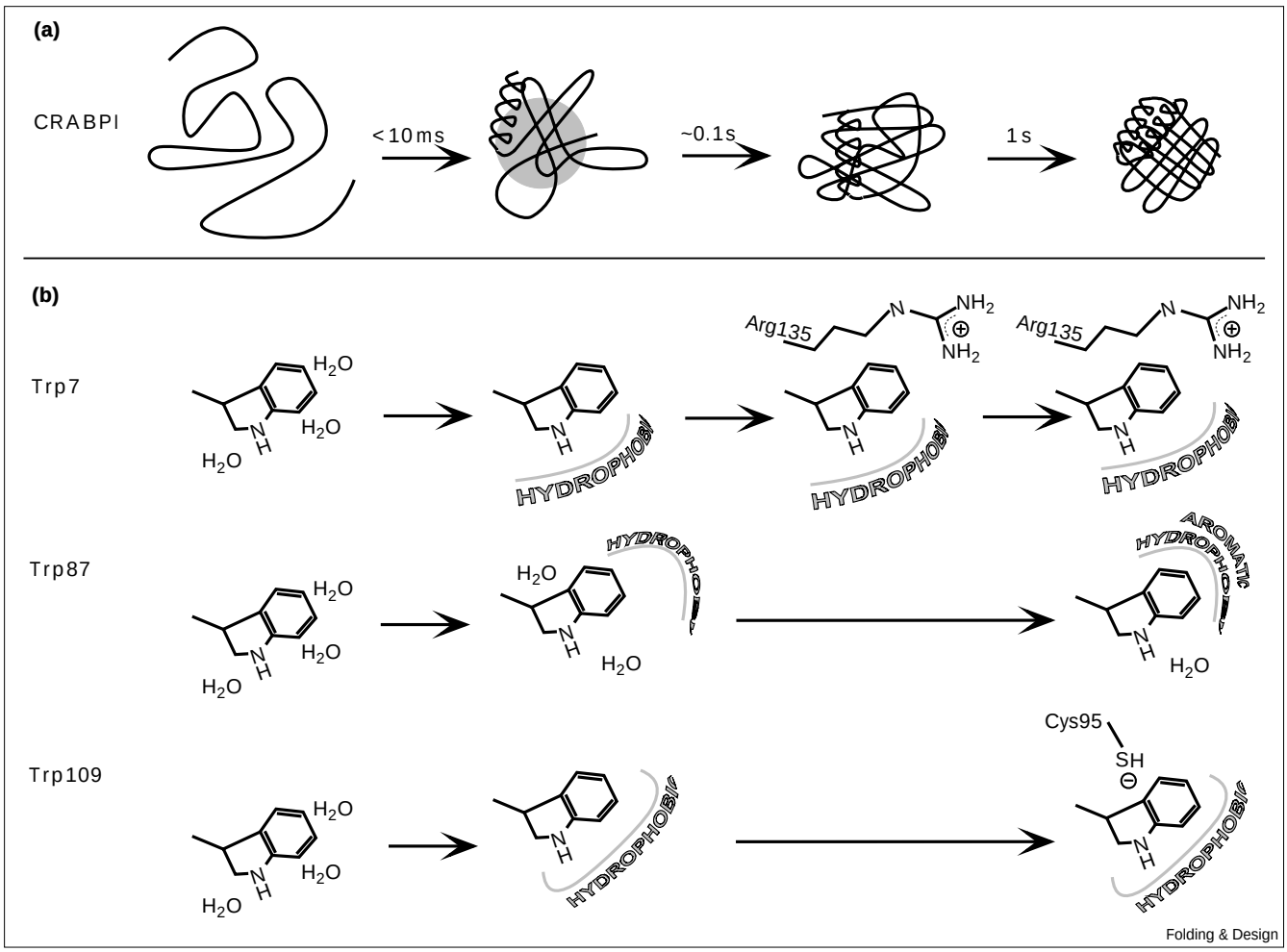
A minor population of CRABPI experiences a final folding step

This final folding step (not shown in Figure 9) has a time constant of > 10 s and arises from CRABPI molecules undergoing isomerization of one or more proline peptide bonds [33,34]. Curiously, this isomerization contributes significantly to the observed fluorescence amplitude for Trp87 (40%) and Trp109 (30%) and therefore must require significant structural rearrangements in the latter half of the protein sequence. There are three proline residues within the CRABPI sequence, at positions 39, 85 and 105 (plus another at the N terminus). Late rearrangements at the Pro85 position have been conclusively shown to be the cause of the > 10 s time constant in the folding of both WT* and Trp87-only (S. Eyles and L.M.G., manuscript in preparation). The proline-derived rearrangements do not have a substantial effect on the environment of Trp7.

This model for CRABPI folding has illuminated several steps in the process of forming the native structure. It is clear that folding is initiated via global hydrophobic collapse, during which substantial structural development occurs. Large amounts of secondary structure are formed during the collapse, indicating that the initial condensation of the chain is a specific structural reaction and not merely a nonspecific burial of hydrophobic sidechains. It is also clear that native structural elements develop first in the proximity of Trp7, on the same time scale as regular β-sheet secondary structure formation (100–200 ms), and ligand-binding cavity formation. These stages are followed by the formation of the remaining specific tertiary interactions (1 s), typified by the Trp109–Cys95 contacts between neighboring strands of the β-sheet network. Concurrent with this fixing of the sidechain rotamers, the stable native hydrogen-bonding network develops cooperatively [25], leading to the final native structure. Hence, the species present after 100–200 ms are most likely to contain a solvated or 'swollen' β sheet with native-like strand register but greater than native strand–strand distances.

Although our current conclusions are drawn from experiments with a single β-sheet topology, they present

Figure 9



A model for the folding pathway of CRABPI, derived from the conclusions of the SF-Flu experiments. (a) General characteristics of the CRABPI folding intermediates, derived from the wild-type folding data. (b) A closer examination of the evolution of the native tertiary environment of each tryptophan residue. The shaded bands and circle represent the presence of hydrophobic surfaces or clusters. See text for details.

intriguing possibilities for the folding pathways of other β-sheet and cavity-containing proteins. Sheet formation follows a concerted hydrophobic collapse; yet this 'collapse' is to an early intermediate ensemble that already contains significant structural components. β-sheet secondary structure forms before the development of specific tertiary interactions, suggesting that secondary structure evolves from the earlier, transient, less stable interactions between amino acid residues. The subsequent evolution of the bulk of the native fluorescence environment of Trp7 indicates that the surrounding residues (including the C terminus) are the sites of early stable native structure formation; similar early docking of the N and C termini is seen during the folding of cytochrome *c* [35,36]. Native tertiary interactions and hydrogen bonds [25] develop more slowly; therefore, while they clearly stabilize the final folded structure of a protein, these interactions do not appear to be the driving force behind β-sheet formation. Future studies on a wide range of simple β-sheet native structures will test whether these findings are specific for CRABPI and its β-clamshell family members, or whether they represent more general mechanisms.

Materials and methods

Construction of CRABPI mutant vectors

To create a His-tagged version of CRABPI/R131Q (WT*), the previously generated pT7-7/CRABP/R131Q vector [22] was digested with *Nde*I and *BamH*I and subcloned into the *E. coli* expression vector pET-16b (Novagen). This vector contains an N-terminal His_{10} tag and a short linker with a Factor Xa cleavage site. To create the single-Trp mutants of WT*, site-directed mutagenesis was performed using the inverse-PCR method [37]. The oligonucleotides (Life Technologies) and templates used, as well as the mutants constructed, are shown in Table 1. Lyophilized primers were dissolved in water to a concentration of 100 pmol/μl and phosphorylated by the following method: the 20 μl reaction solution (5 pmol/μl primer, 1× kinase buffer (50 mM Tris

pH 7.5, 10 mM $MgCl_2$, 5 mM dithiothreitol (DTT), 50 μg/ml BSA), 1 mM rATP, 10 units of T4 kinase (Gibco BRL)) was incubated at 37°C for 1 h, followed by 70°C for 10 min, then chilled on ice to room temperature. PCR reactions (25 μl) consisted of: 0.4 pmol/μl of each phosphorylated primer, 0.2 mM dNTP mix (Gibco BRL), 2 ng template DNA, 2.5 units *Pfu*I polymerase (Stratagene), and 1× *Pfu*I buffer, and were subject to the following program: denaturation (94°C, 4 min), amplification 25× ((94°C, 1 min) (60°C, 1 min) (72°C, 12 min)), and extension (72°C, 30 min). PCR products were purified on a 1% agarose gel, with a band corresponding to full-length plasmid (6.4 kb) present as the only significant product. DNA from these bands was extracted with a QiaQuick kit (QIAGEN) and resuspended in 50 μl of TE buffer (10 mM Tris pH 7.5, 1 mM EDTA pH 8.0). Half of the purified PCR product was ligated overnight at 16°C, and half of the ligation solution was used to transform *E. coli* strain DH5α, which was plated onto LB agar containing 0.1 mg/ml ampicillin. These conditions typically resulted in 5–10 transformants, of which approximately 70% were shown (by DNA sequencing) to have incorporated the correct mutation and have maintained fidelity throughout the rest of the gene sequence. Because the mutagenesis was carried out using the pET-16b/WT* expression vector as a template, no additional subcloning steps were necessary.

Protein purification

Expression plasmids containing WT* and mutant CRABPI vectors were used to transform *E. coli* strain BL21(DE3) [38]. Cell cultures were grown and harvested as described [24]. Cell pellets from 1.5 l of culture were resuspended in 45 ml of SL buffer (50 mM sodium phosphate pH 8.0 and 300 mM NaCl), flash frozen in liquid nitrogen, and stored at –80°C. Thawed cells were lysed by addition of 25 mg lysozyme (Sigma) and incubated on ice for 30 min. Ten minutes of sonication (20 s bursts with 100 s intervals between bursts) was used to shear the DNA. CRABP-WT* was expressed primarily in the cytoplasm, while all single-Trp CRABPI mutant proteins were expressed as part of the insoluble cell fraction. The cell suspension was centrifuged for 30 min at 27000*g*, and either the supernatant (WT*) or pellet (Trp-mutant proteins) was subjected to additional purification steps.

For WT*, the supernatant was filtered through a sterile 0.45 μm syringe filter and applied to a 10 ml Ni-NTA (QIAGEN) column. The loaded column was washed with 150 ml W buffer (20 mM imidazole, 10 mM Tris pH 8.0 and 0.5 M NaCl) and eluted with a linear gradient of 20–500 mM imidazole (WT* eluted at ~200 mM imidazole). Fractions containing WT* were pooled and judged > 90% pure by Coomassie-stained SDS–PAGE. Pooled fractions were dialyzed extensively against either 10 mM Tris pH 8.0 or Milli-Q water (Milligen), and either stored at 4°C for use within two weeks, or lyophilized.

For Trp-mutant proteins, pellets were resuspended in 20 ml U buffer (8 M urea, 5 mM imidazole, 0.5 M NaCl) and centrifuged at 27000*g* for 1 h. The supernatant was filtered through a sterile 0.45 μm syringe filter and applied to a 10 ml Ni-NTA (QIAGEN) column, equilibrated with U buffer. The loaded column was washed with U buffer and eluted with a linear gradient of 5–500 mM imidazole. Fractions were pooled and analyzed as for WT* preparation. Pooled fractions were treated with β-mercaptoethanol (βME) to a final concentration of 100 mM. Proteins were refolded by dripping the protein/urea solution from a 10 ml disposable syringe (plunger removed) fitted with a 0.45 μm syringe filter into 4 l of rapidly stirred refolding buffer (10 mM Tris pH 8.0, 10 mM βME) at 4°C; the syringe and filter regulated the rate of protein/urea addition such that each drop was completely mixed with the refolding buffer before addition of the next drop. After addition was complete, the solution was allowed to stir overnight and then centrifuged at 23000*g* for 1 h to remove misfolded aggregates. The clarified protein solutions were concentrated to approximately 1 mg/ml using an Amicon concentrator and dialyzed and stored as described for WT*.

Equilibrium CD and fluorescence measurements

Far-UV CD spectra were collected on a Jasco J-715 spectropolarimeter, using a cell with a 2 mm path length and a 1 nm bandwidth. Steady-state fluorescence measurements were made using a PTI QM-1 spectrofluorimeter, with excitation at 280 nm (1.5 nm slit width) and emission from 300–380 nm (3 nm slit width). All fluorescence and CD spectra were collected at 25°C and corrected by subtracting a solvent spectrum acquired under identical conditions. Protein concentrations and buffer conditions are described in the corresponding figure legends.

Urea titrations

A series of samples was made containing 0–8 M urea, all with 6.7 μM protein (diluted from a stock in 10 mM Tris pH 8.0) and 10 mM Tris pH 8.0. Samples were mixed well and allowed to unfold for ≥ 2 h. Final conditions for urea refolding samples were identical, except that 100 mM βME was included. Refolding samples were made by diluting the protein stock solution (in 10 M urea with 10 mM Tris pH 8.0 and 100 mM βME) into refolding buffer, with vigorous stirring, to a variety of final urea concentrations. Refolding samples were allowed to equilibrate for ≥ 2 h before fluorescence measurements. For all urea titrations, fluorescence spectra were acquired as described above.

Retinoic acid binding

The ability of mutant proteins to bind retinoic acid was assayed by the fluorescence quenching method of Cogan *et al.* [28]. WT* and Trp-mutant proteins (1 μM in 10 mM Tris pH 8.0) were excited at 280 nm and their fluorescence spectra measured from 300–380 nm. Retinoic acid (RA), prepared in absolute ethanol at several stock concentrations, was added such that the final added volume did not exceed 4% of the protein solution. RA stock concentrations were determined from the absorbance at 336 nm (ε = 45,000 M^{-1} cm^{-1} [28]).

Stopped-flow fluorescence

Sample preparation was as described in [26], except that 100 mM βME was included in the folding buffer and unfolded protein stock. The presence of βME did not alter the folding kinetics of WT*, but did have an effect on the kinetics of the mutant proteins and was necessary for complete recovery of the native fluorescence spectrum of Trp87-only. Presumably, the addition of βME during the folding of Trp87-only destabilizes a transiently disulfide-bonded intermediate that is not populated during the folding of WT*. Final sample conditions were 0.08 mg/ml protein refolding in 10 mM Tris pH 8.0, 100 mM βME with 0.54 M urea. Unfolding data were acquired under identical shot conditions using refolding buffers by diluting the protein from 10 mM Tris pH 8.0, 100 mM βME (with 2 M urea for WT*) into various concentrations of urea.

Kinetic data were collected and processed as previously described [26]. Kinetic constants for each trace were obtained by fitting to a summed exponential equation, with the quality of fit determined by the randomness of residuals. The fluorescence spectrum of the unfolded protein, taken in 7 M urea, was corrected, based on the urea dependence of the fluorescence spectrum (see Figure 4 and [39]), to correspond to the spectrum of unfolded protein in 0.54 M urea [39].

Stopped-flow CD

Kinetic CD data were collected via a Bio-Logic SFM-4 stopped-flow apparatus (Claix, France) attached to a Jasco 715 spectropolarimeter. Analog signals for both uncorrected CD and PMT voltages were routed to the Bio-Logic BioKine software for collection. The solution conditions, dilution ratio, and syringe speeds were identical to those used for SF-Flu experiments [26]. For measuring the dead-time CD spectrum, the response time on the CD instrument was 125 ms (corresponds to a time constant of 15 ms), and the bandwidth was set to 2 nm. At each wavelength, data were collected for 1 s after the initiation of each shot, with a time constant of 100 ms for each point (that is, 11 points/shot). Data from two separate shots were averaged, and the voltages from 0.2–0.8 s were then averaged. Complete equilibrium spectra were collected for refolded CRABPI and unfolded CRABPI (in 7 M urea, 10 mM Tris pH 8.0), using a response time of 4 s and a scan speed of 20 nm/min; five scans were averaged. Both the equilibrium spectra and

the averaged continuous push spectrum were corrected for solvent and cuvette contributions by subtracting a corresponding buffer blank spectrum.

For single-wavelength folding kinetics, the CD instrument was set for a response time of 8 ms (time constant of 2 ms) and a bandwidth of 2 nm, and data were collected for 20 s after each shot, using a split time base (10 ms time constant for 1.4 s, 20 ms time constant for remainder). Approximately 200 shots were averaged for each wavelength. The averaged kinetic traces were smoothed, using a five-point base, within the BioKine software package, and converted to ellipticity.

Acknowledgements

The authors would like to thank Steve Eyles, Phil Thomas, Zhi-Ping Liu, Ed Feng and Bruce Hudson for helpful discussions, and Rolf Fuerter and Jeff Clark for advice about expression systems. This research was supported by NIH grant GM27616. P.L.C. was supported by the NIH Biophysics Predoctoral Training Program (GM08237).

References

1. Udgaonkar, J.B. & Baldwin, R.L. (1990). Early folding intermediate of ribonuclease A. *Proc. Natl Acad. Sci. USA* **87**, 8197-8201.
2. Dobson, C.M., Evans, P.A. & Radford, S.E. (1994). Understanding how proteins fold: the lysozyme story so far. *Trends Biochem. Sci.* **19**, 31-37.
3. Fersht, A.R. (1993). Protein folding and stability: the pathway of folding of barnase. *FEBS Lett.* **325**, 5-16.
4. Scholtz, J., Barrick, D., York, E., Stewart, J. & Baldwin, R. (1995). Urea unfolding of peptide helices as a model for interpreting protein unfolding. *Proc. Natl Acad. Sci. USA* **92**, 185-189.
5. Rose, G.D. & Wolfenden, R. (1993). Hydrogen bonding, hydrophobicity, packing, and protein folding. *Annu. Rev. Biophys. Biomol. Struct.* **22**, 381-415.
6. Capaldi, A.P. & Radford, S.E. (1998). Kinetic studies of β-sheet protein folding. *Curr. Opin. Struct. Biol.* **8**, 86-92.
7. Minor, D.L., Jr & Kim, P.S. (1994). Measuring the β-sheet forming propensities of the amino acids. *Nature* **367**, 660-663.
8. Kim, C.A. & Berg, J.M. (1993). Thermodynamic β-sheet propensities measured using a zinc-finger host peptide. *Nature* **362**, 267-270.
9. Dalal, S., Balasubraminian, S. & Regan, L. (1997). Protein alchemy: changing β-sheet into α-helix. *Nat. Struct. Biol.* **4**, 548-552.
10. Parker, M.J., Dempsey, C.E., Lorch, M. & Clarke, A.R. (1997). Acquisition of native β-strand topology during the rapid collapse phase of protein folding. *Biochemistry* **36**, 13396-13405.
11. Schonbrunner, N., Koller, K.P. & Kiefhaber, T. (1997). Folding of the disulfide-bonded β-sheet protein tendamistat: rapid two-state folding without hydrophobic collapse. *J. Mol. Biol.* **268**, 526-38.
12. Agashe, V.R., Shastry, M.C.R. & Udgaonkar, J.B. (1995). Initial hydrophobic collapse in the folding of barstar. *Nature* **377**, 754-757.
13. Houry, W.A. & Scheraga, H.A. (1996). Structure of a hydrophobically collapsed intermediate on the conformational folding pathway of ribonuclease A probed by hydrogen-deuterium exchange. *Biochemistry* **35**, 11734-11746.
14. Engelhard, M. & Evans, P.A. (1995). Kinetics of interaction of partially folded proteins with a hydrophobic dye: evidence that molten globule character is maximal in early folding intermediates. *Protein Sci.* **4**, 1553-1562.
15. Hamada, D., Segawa, S.I. & Goto, Y. (1996). Non-native α-helical intermediate in the refolding of β-lactoglobulin, a predominantly β-sheet protein. *Nat. Struct. Biol.* **3**, 868-873.
16. Hamada, D. & Goto, Y. (1997). The equilibrium intermediate of β-lactoglobulin with non-native α-helical structure. *J. Mol. Biol.* **269**, 479-487.
17. Graciani, N.R., Tsang, K.Y., McCutchen, S.L. & Kelly, J.W. (1994). Amino acids that specify structure through hydrophobic clustering and histidine-aromatic interactions lead to biologically active peptidomimetics. *Bioorg. Med. Chem.* **2**, 999-1006.
18. Engelhard, M. & Evans, P.A. (1996). Experimental investigation of sidechain interactions in early folding intermediates. *Fold. Des.* **1**, R31-R37.
19. Kleywegt, G.J., Bergfors, T., Senn, H., Le Motte, P., Gsell, B., Shudo, K. & Jones, T.A. (1994). Crystal structures of cellular retinoic acid binding proteins I and II in complex with all-trans-retinoic acid and a synthetic retinoid. *Structure* **2**, 1241-1258.
20. Thompson, J., Bratt, J. & Banaszak, L. (1995). Crystal structure of cellular retinoic acid binding protein I shows increased access to the binding cavity due to formation of an intermolecular β-sheet. *J. Mol. Biol.* **252**, 433-446.
21. Banaszak, L., Winter, N., Xu, Z., Bernlohr, D.A., Cowan, S. & Jones, T.A. (1994). Lipid-binding proteins: a family of fatty acid and retinoid transport proteins. *Adv. Protein Chem.* **45**, 89-151.
22. Zhang, Z., Liu, Z.P., Jones, T.A., Gierasch, L.M. & Sambrook, J.F. (1992). Mutating the charged residues in the binding pocket of cellular retinoic acid-binding protein simultaneously reduces its binding affinity to retinoic acid and increases its thermostability. *Proteins* **13**, 87-99.
23. Frieden, C., Jiang, N. & Cistola, D.P. (1995). Intestinal fatty acid binding protein: folding of fluorescein-modified proteins. *Biochemistry* **34**, 2724-2730.
24. Liu, Z.P., Rizo, J. & Gierasch, L.M. (1994). Equilibrium folding studies of cellular retinoic acid binding protein, a predominantly β-sheet protein. *Biochemistry* **33**, 134-142.
25. Clark, P.L., Liu, Z.P., Rizo, J. & Gierasch L.M. (1997). Cavity formation before stable hydrogen bonding in the folding of a β-clam protein. *Nat. Struct. Biol.* **4**, 883-886.
26. Clark, P.L., Liu, Z.P., Zhang, J. & Gierasch, L.M. (1996). Intrinsic tryptophans of CRABPI as probes of structure and folding. *Protein Sci.* **5**, 1108-1117.
27. Greenfield, N. & Fasman, G.D. (1969). Computed circular dichroism spectra for the evaluation of protein conformation. *Biochemistry* **8**, 4108-4116.
28. Cogan, U., Kopelman, M., Mokady, S. & Shinitzky, M. (1976). Binding affinities of retinol and related compounds to retinol binding proteins. *Eur. J. Biochem.* **65**, 71-78.
29. Creighton, T.E. (1990). Protein folding. *Biochem. J.* **270**, 1-16.
30. Lakowicz, J.R. (1983). *Principles of Fluorescence Spectroscopy*. Plenum Press, New York, NY.
31. Freifelder, D. (1982). *Physical Biochemistry*. (2nd edn), W.H. Freeman and Company, New York, NY.
32. Sukumar, M. & Gierasch, L.M. (1997). Local interactions in a Schellman motif dictate interhelical arrangement in a protein fragment. *Fold. Des.* **2**, 211-222.
33. Walkenhorst, W.F., Green, S.M. & Roder, H. (1997). Kinetic evidence for folding and unfolding intermediates in staphylococcal nuclease. *Biochemistry* **36**, 5795-5805.
34. Koide, S., Dyson, H.J. & Wright, P.E. (1993). Characterization of a folding intermediate of apoplastocyanin trapped by proline isomerization. *Biochemistry* **32**, 12299-12310.
35. Bai, Y., Sosnick, T.R., Mayne, L. & Englander, S.W. (1995). Protein folding intermediates: native state hydrogen exchange. *Science* **269**, 192-196.
36. Sosnick, T.R., Mayne, L. & Englander, S.W. (1996). Molecular collapse: the rate-limiting step in two-state cytochrome *c* folding. *Proteins* **24**, 413-26.
37. Hemsley, A., Arnheim, N., Toney, M., Cortopassi, G. & Galas, D. (1989). A simple method for site-directed mutagenesis using the polymerase chain reaction. *Nucleic Acids Res.* **17**, 6545-6551.
38. Studier, F.W., Rosenberg, A.H., Dunn, J.J. & Dubendorff, J.W. (1990). Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol.* **185**, 60-89.
39. Pace, C.N., Shirley, B.A. & Thomson, J.A. (1989). Measuring the conformational stability of a protein. In *Protein Structure: A Practical Approach*. (Creighton, T.E., ed.), 311-330, Oxford University Press, UK.