doi:10.1006/jmbi.2001.4436 available online at http://www.idealibrary.com on IDEAL® *J. Mol. Biol.* (2001) **306**, 877–888

JMB

Native Hydrogen Bonds in a Molten Globule: The Apoflavodoxin Thermal Intermediate

María P. Irún[1], Maria M. Garcia-Mira[2], Jose M. Sanchez-Ruiz[2] and Javier Sancho[1]*

[1]*Departamento de Bioquímica y Biología Molecular y Celular Facultad de Ciencias Universidad de Zaragoza 50009 Zaragoza, Spain*

[2]*Departamento de Química Física. Facultad de Ciencias Universidad de Granada 18071 Granada, Spain*

**Corresponding author*

The structure and energetics of protein-folding intermediates are poorly understood. We have identified, in the thermal unfolding of the apoflavodoxin from *Anabaena* PCC 7119, an equilibrium intermediate with spectroscopic properties of a molten globule and substantial enthalpy and heat capacity of unfolding. The structure of the intermediate is probed by mutagenesis (and phi analysis) of polar residues involved in surface-exposed hydrogen bonds connecting secondary-structure elements in the native protein. All hydrogen bonds analysed are formed in the molten globule intermediate, either with native strength or debilitated. This suggests the overall intermediate's topology and surface tertiary interactions are close to native, and indicates that hydrogen bonding may contribute significantly to shape the conformation and energetics of folding intermediates.

Keywords: protein stability; hydrogen bond; molten globule; protein intermediate; protein folding

Introduction

The stability of native proteins, relative to the unfolded state, has been actively investigated[1-5] and can presently be increased in a rational manner.[6–13] In addition to the native and unfolded states, proteins can adopt, either transiently[14] or at equilibrium,[15] other conformations that display mixed properties of the native and denatured states. The role of transient folding intermediates is controversial as they could, in principle, help accelerate the reaction or rather act as kinetic traps.[14,16] As for the equilibrium intermediates, they have in some cases been identified with the folding intermediates[17–19] but, in other cases, they are simply regarded as unwanted conformations that may contribute to aggravate aggregation phenomena within the cell.[20] Equilibrium intermediates can be detected as a deviation from two-state behaviour (i.e. non-coincident unfolding curves obtained with different techniques). Compared to those of the native state, the structure and energetics of protein intermediates are poorly understood. A number of mutational studies on these intermediates have focused on apolar/apolar packing side-chain interactions,[21–29] while polar interactions, such as hydrogen bonds, have received much less attention.[30] To investigate the role of side-chain hydrogen bonds in the structure of protein intermediates, we have characterised the energetics of an apoflavodoxin[31-33] thermal intermediate and then probed the presence of native hydrogen bonds by performing a mutational analysis of the three-state unfolding equilibrium. All hydrogen bonds analysed here are shown to be formed in the intermediate, although most are debilitated relative to the native state.

Abbreviations used: DSC, differential scanning calorimetry; C_p, heat capacity; WT, wild-type; N, native; I, intermediate; U, unfolded.

E-mail address of the corresponding author: jsancho@posta.unizar.es

Results and Discussion

An equilibrium intermediate in the thermal unfolding of *Anabaena* apoflavodoxin

Flavodoxins are α/β-electron transfer flavoproteins that carry one molecule of non-covalently bound flavin mononucleotide that can be reversibly removed to yield the apoflavodoxin. We use the apoflavodoxin from *Anabaena*[31–33] to study protein stability and binding.[33–37] The equilibrium chemical unfolding of the apoflavodoxin from *Anabaena* follows a two-state mechanism;[34] yet an intermediate appears in the folding reaction

0022-2836/01/040877–12 $35.00/0

(Fernández-Recio *et al.*, unpublished results) and molten globule conformations are dominant at low pH[34] and in protein variants whose stability has been reduced by truncation.[35] Since thermal unfolding may reveal equilibrium intermediates that are not observed by chemical denaturation, and since an intermediate has been reported in the thermal unfolding of the apoflavodoxin from *Azotobacter vinelandii*,[38] we have investigated the possible occurrence of intermediate conformations in the thermal unfolding of the *Anabaena* apoflavodoxin by performing the so-called superposition test.[29,35] To that end, we have compared the unfolding curves obtained using four different spectroscopic techniques: tryptophan fluorescence emission, far-UV circular dichroism (CD), near-UV absorption, and near-UV CD (the near-UV absorption and CD curves being obtained simultaneously). In all cases, the native signal was recovered upon cooling, and no protein concentration effect on T_m was observed in the concentration range used (not shown). The unfolding curves obtained for wild-type apoflavodoxin are shown in Figure 1. Both the fluorescence and near-UV CD curves, and the absorbance and far-UV curves can be superimposed reasonably. However, the two sets of curves are clearly non-superimposable. Simple two-state fits of the individual curves (shown in Figure 1) indicate that the apparent melting temperatures obtained by fluorescence and near-UV CD are around 7 K lower than those from the two other techniques. Besides, the unfolding enthalpies derived from the fluorescence and near-UV CD curves are of around 25 kcal mol^{-1}, while those from the absorbance and far-UV CD curves are around 50 kcal mol^{-1}. These data clearly indicate that the thermal unfolding of the protein is not two-state and that an intermediate accumulates in the transition region.

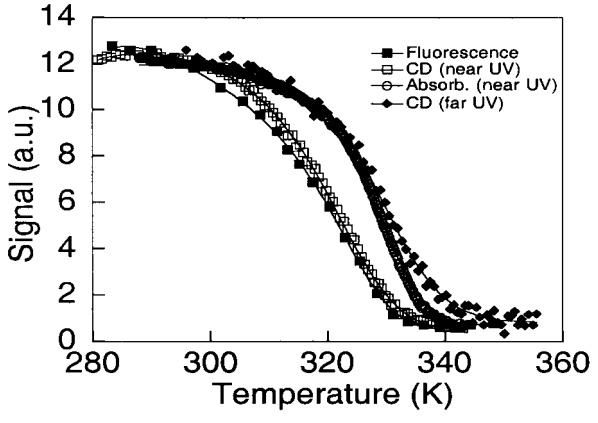

Figure 1. Thermal unfolding of wild-type apoflavodoxin followed by fluorescence, near-UV absorbance, and far-UV and near-UV circular dichroism. Fluorescence conditions: 2 μM apoflavodoxin in 50 mM Mops (pH 7.0); ratio of emission at 320 and 360 nm. Near-UV absorbance and circular dichroism conditions (curves obtained simultaneously): 40 μM apoflavodoxin in 50 mM Mops (pH 7.0); signal recorded at 291 nm. Far-UV circular dichroism conditions: 1 μM apoflavodoxin in 5 mM Mops (pH 7.0) with 15 mM NaCl; ellipticity at 222 nm. The continuous lines are fits to a simple two-state model.

The above conclusion would seem to conflict with a previous calorimetric study[34] in which the differential scanning calorimetry (DSC) transitions for apoflavodoxin unfolding could be described adequately by the two-state model. The discrepancy is only apparent, however, since the enthalpy change associated with the native to intermediate step (about 25 kcal mol^{-1}) is comparatively small and would give rise to a very broad calorimetric transition that could easily be "lost" in the instrumental baseline uncertainties.[39] In this work, we take advantage of the high sensitivity and outstanding baseline reproducibility of the VP-DSC calorimeter[40] to obtain very accurate heat capacity data for apoflavodoxin unfolding in a wide temperature range. Figure 2 shows the raw calorimetric data for a representative DSC experiment where six successive buffer-buffer baselines were taken before, and one after, running the protein sample; note the outstanding baseline reproducibility. The absolute heat capacity values obtained from the raw data in Figure 2 are given in Figure 3. The upper panel shows that the fitting to the two-state model is, in principle, reasonable. However, a significant and systematic deviation is observed at the lower temperatures and, more important, the calculated temperature-dependencies of the heat capacity for the native (C_p(N)) and unfolded (C_p(U))states cross at about 70 °C. This clearly unrealistic result argues against the two-state model, as it would imply that the unfolding heat capacity change became negative above that temperature. We have then fitted the data to a three-state model, using a parameter (Z) to describe the similarity of the heat capacity of the

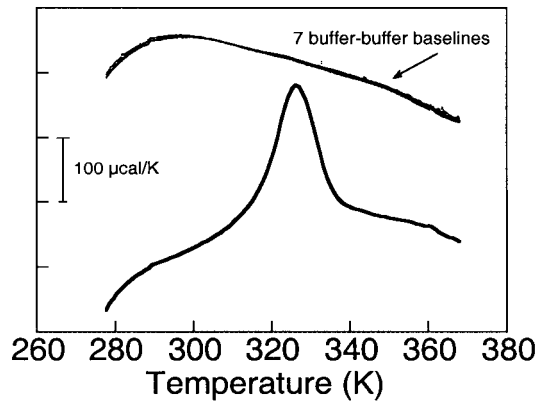

Figure 2. Original DSC thermogram for a solution of apoflavodoxin (concentration 1.82 mg ml^{-1}; scan rate 0.5 K min^{-1}) and the corresponding buffer-buffer baselines. Six buffer-buffer baselines were obtained before the protein run, and one additional buffer-buffer baseline was recorded afterwards.

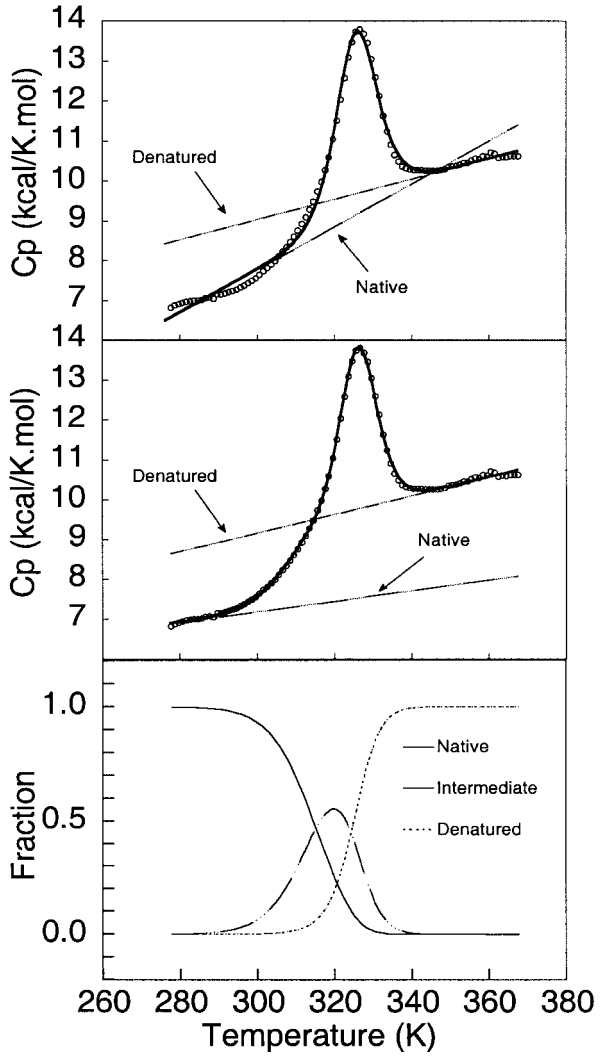

Figure 3. Upper panel: absolute heat capacity (open circles) *versus* temperature profile for apoflavodoxin unfolding derived from the raw data shown in Figure 2; the continuous line is the best fit to the two-state model and the broken lines represent the calculated heat capacities for the native and unfolded states. Middle panel: same as in the upper panel but here the continuous line is the best fit to the three-state model (with a value of 0 for the Z parameter (see Materials and Methods for details)). Lower panel: population of states *versus* temperature corresponding to the three-state fit shown in the middle panel.

intermediate state to those of the native and unfolded states (if $C_p(I) = C_p(U)$, $Z = 1$ and if $C_p(I) = C_p(N)$, $Z = 0$, where $C_p(I)$ is the heat capacity of the intermediate state). We find excellent fits to the entire curve (see the middle panel in Figure 3 for an example) and, for Z values below 0.6, reasonable temperature dependencies for $C_p(N)$ and $C_p(U)$, which no longer cross each other. Similar results were obtained from the analysis of four additional DSC profiles recorded at different protein concentrations and scanning rates; the parameters derived from these three-state fits are shown in Table 1.

Thermodynamic and spectroscopic properties of the intermediate

The availability of four non-superimposable unfolding curves (Figure 1) allows a global analysis of the spectroscopic data to a three-state equilibrium unfolding mechanism. We have followed the approach described by Luo *et al.*[41] where all thermodynamic properties of the two equilibria (native/intermediate and intermediate/unfolded) are globally constrained but the degree of spectroscopic resemblance between the intermediate and the denatured state (the Z parameter, see Materials and Methods) is allowed to vary among the four techniques used to probe the unfolding. The global fit of the four curves is shown in Figure 4, upper panel. The melting temperatures of the two equilibria are well separated (316.5 and 329.6 K), the maximal accumulation of intermediate (64%) occurs at 323 K (Figure 4, lower panel), and the

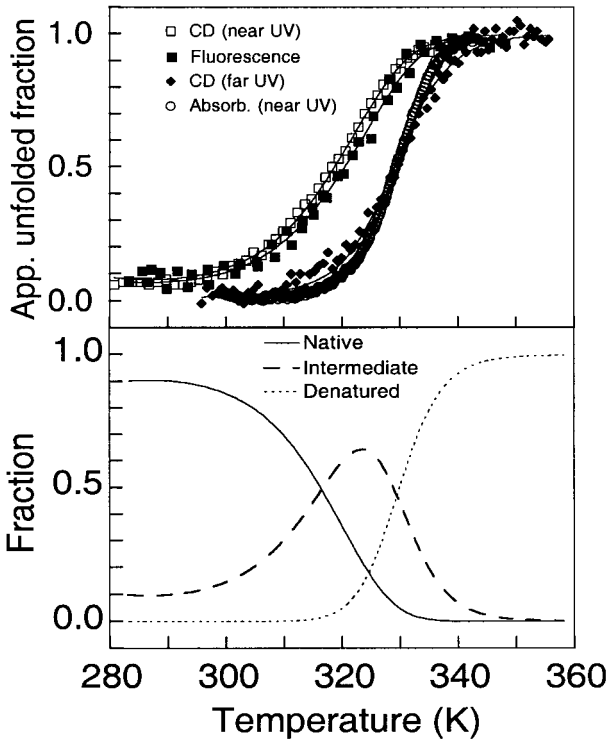

Figure 4. Upper panel: global fit of the fluorescence, near-UV absorbance, and near and far-UV circular dichroism thermal unfolding curves of wild-type apoflavodoxin to a three-state model. Data are shown as apparent unfolded fractions. The continuous lines are the best fit, with all thermodynamic parameters being the same for the four curves. The spectroscopic similarities of the intermediate and unfolded states are allowed to vary among the curves. Lower panel: population of states *versus* temperature corresponding to the three-state fit shown in the upper panel.

Table 1. Thermodynamic data of the NI and IU equilibria in the apoflavodoxin thermal unfolding and spectroscopic and calorimetric resemblance between the intermediate and unfolded and native states

	T_{mNI} (K)	ΔH_{NI} (kcal mol^{-1})	T_{mIU} (K)	ΔH_{IU} (kcal mol^{-1})	$\Delta C_{p,\,NU}$ (kcal mol^{-1} K^{-1})	Z[c]				
						Fluor.	CD near-UV	CD Far-UV	Absor.	Calor.
Global fit[a] (spectroscopic)	316.5 ± 0.3	28.1 ± 0.5	329.6 ± 0.1	54.9 ± 1.2	-	0.7	0.8.	0.1	0.1	-
Calorimetric fit[b]	311.8 ± 2.5	28.0 ± 3.7	326.5 ± 1.2	56.0 ± 2.5	2.5 ± 0.3	-	-	-	-	<0.6[d]

[a] Global fit of the thermal unfolding curves of wild-type apoflavodoxin, followed by fluorescence, near-UV absorbance and near and far-UV circular dichroism, to a three-state model (Figure 4, upper panel). Two curves of each technique were included in the global fit. All thermodynamic parameters of the two equilibria are kept identical for all the curves while the spectroscopic similarity between the intermediate and the unfolded states (Z parameters) can vary among the different techniques. Errors are provided by the fitting program.

[b] Fit of calorimetric thermograms to a three-state model (Figure 2, middle panel). The data shown are averages (±SD) of the fits corresponding to five different thermograms recorded at protein concentrations between 1.1 and 4.6 mg ml^{-1} and at scan-rates between 0.5 and 1.5 K min^{-1}. No protein concentration effect or scan rate effect was noticed. Each thermogram was first independently fitted to the three-state model using several different Z values (below 0.6). All the fits of a given thermogram, performed using different Z values, gave similar values for the transition temperatures and enthalpy changes. The calculated ΔC_p corresponds to the total unfolding heat capacity change. The physical meaning of Z is different in the spectroscopic and in the calorimetric fit (see Materials and Methods for details).

[c] Similarity of the intermediate to the unfolded and native states, as defined in the methods. $Z = 1$ if the intermediate is identical with the unfolded state and $Z = 0$ if the intermediate is identical with the native state.

[d] The calorimetric similarity refers to the heat capacity (see Materials and Methods). Any value below 0.6 is compatible with the available calorimetric unfolding data.

enthalpies are 28 and 55 kcal mol^{-1}, respectively. These results are in very good agreement with the calorimetric parameters derived from the three-state fit (Table 1). According to these data, the first transition (native to intermediate) involves a debilitation of the interactions contributing to the enthalpic stabilisation of the native protein to roughly two-thirds of their strength in the native state. The intermediate, thus, still displays a substantial enthalpic stabilisation relative to the unfolded state. The distribution of species at different temperatures predicted by the global analysis of the spectroscopic data (Figure 4, lower panel) is similar to that derived from the calorimetric analysis (Figure 3, lower panel). At low temperature, the spectroscopic fit predicts a non-negligible intermediate concentration, which is also predicted by the calorimetric analysis at certain Z values (not shown). This prediction remains to be tested.

Analysis of the DSC profiles reported in this work provides an estimate for the total heat capacity change for apoflavodoxin unfolding (see Table 1), but not for the heat capacity of the intermediate state relative to the native or unfolded states, since good fittings to the three-state model could be obtained with different values of Z (see Materials and Methods for details). To estimate the heat capacity of the intermediate state, we have resorted to an analysis of the temperature-dependence of the transition enthalpies (Table 2) for wild-type (WT) apoflavodoxin and the several mutants that have been used in this work to investigate the topology of the intermediate (see below). The rationale is that $\Delta\Delta H$ and $\Delta\Delta C_p$ values (WT minus mutant) for mutations that do not alter the native structure (or the average structure of the denatured ensemble) may be expected to be negligible compared with ΔH and ΔC_p. We thus attribute the rather large variation observed in the experimental ΔH values in Table 2 to the heat capacity effect, since the transition ΔH values correspond to different transition temperatures. With this assumption, the slope of a plot of transition ΔH *versus* T_m from the different protein variants should provide an estimate of ΔC_p. We show in Figure 5 two such plots corresponding to the N-to-I transition and for the global unfolding process, N-to-U (i.e. $\Delta H_{NI} + \Delta H_{IU}$ *versus* the average of T_{mNI} and T_{mIU}). From the slope of the latter plot, we obtain a value of 3.1(±0.8) kcal K^{-1} mol^{-1} for the global unfolding heat capacity change, in good agreement with the value derived from the analysis of the calorimetric data (2.5(±0.3) kcal K^{-1} mol^{-1}; Table 1). Similarly, the heat capacity change associated with the N-to-I transition is estimated to be 1.5(±0.3) kcal K^{-1} mol^{-1}, which indicates that significant surface burial takes place in the intermediate.

As for the spectroscopic properties of the intermediate, inferred from the values of the spectroscopic Z parameters (Table 1), our analysis indicates that the intermediate fluorescence and near-UV CD are close to those of the unfolded state, while the absorbance and far-UV CD are almost identical with those of the native protein. This is precisely the expected spectroscopic behaviour of molten globule conformations, which are rich in secondary structure but usually devoid of near-UV signal as a consequence of the debilitation of tertiary interactions. Spectroscopically, the apoflavodoxin thermal unfolding intermediate is thus a molten globule, and thermodynamically, a debilitated version of the native state with substantial enthalpic stabilisation, substantial entropic destabilisation, and substantial heat capacity change. This is in qualitative agreement with the properties of the apoflavodoxin molten globule that appears, at room temperature, upon dissection of the C-terminal helix.[35]

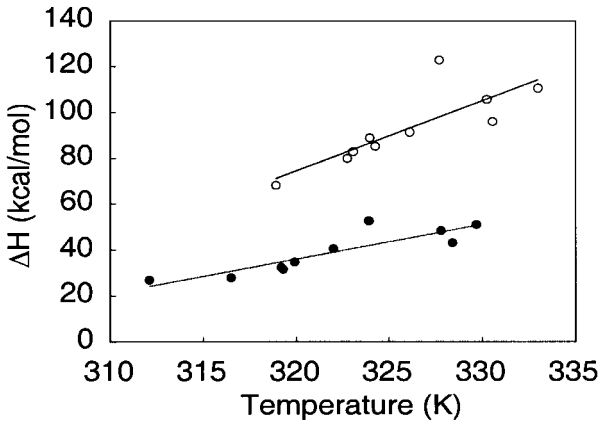

Figure 5. Plots of transition enthalpy *versus* temperature for the N to I transition (filled symbols) and the N to U transition (open symbols), including data for wild-type apoflavodoxin and the several mutants in Table 2. Transition enthalpies for the N to I step (ΔH_{NI}) are assigned to the corresponding transition temperatures (T_{mNI}). Transition enthalpies for the N to U step are calculated as $\Delta H_{NI} + \Delta H_{IU}$ and assigned to the average of the two transition temperatures (that is to $(T_{mNI} + T_{mIU})/2$).

Hydrogen bonding in the intermediate relative to the native state

The influence of side-chain interactions on protein energetics can be analysed by comparison of wild-type and mutant protein stabilities. Although much information is available on the energetics of native states, little is known concerning intermediates. We use here the apoflavodoxin thermal intermediate as a model to probe the relevance of hydrogen bonds in shaping the structure and energetics of protein intermediates. To that end, we have selected, and altered by mutation, eight solvent-exposed hydrogen bonds (Figure 6, Table 3). The thermal unfolding of each mutant has been recorded using the same four spectroscopic techniques used to study the wild-type pro-

Table 2. Thermodynamic parameters of the three-state unfolding equilibrium of wild-type and mutant apoflavodoxins

Protein	ΔH_{NI} (kcal mol^{-1})	T_{mNI} (K)	ΔS_{NI} (cal mol^{-1} K^{-1})	$\Delta\Delta G_{NI}$ (kcal mol^{-1})	ΔH_{IU} (kcal mol^{-1})	T_{mIU} (K)	ΔS_{IU} (cal mol^{-1} K^{-1})	$\Delta\Delta G_{IU}$ (kcal mol^{-1})	$\Delta\Delta G_{NU}$ (kcal mol^{-1})
Wild-type	28.1 ± 0.6	316.5 ± 0.3	89 ± 2	-	54.9 ± 1.2	329.6 ± 0.1	166 ± 4	-	-
Tyr47Phe	26.9 ± 0.5	312.1 ± 0.2	86 ± 2	0.39 ± 0.03	41.4 ± 0.9	325.7 ± 0.1	127 ± 3	0.65 ± 0.03	1.04 ± 0.04
Glu67Ala	32.5 ± 0.8	319.2 ± 0.4	102 ± 2	−0.24 ± 0.04	56.7 ± 1.5	328.7 ± 0.1	173 ± 5	0.15 ± 0.02	−0.09 ± 0.05
pWild-type[a]	31.8 ± 2.3	319.3 ± 1.0	100 ± 7	-	53.7 ± 1.9	329.2 ± 0.3	163 ± 6	-	-
Glu16Gln[a]	51.3 ± 2.3	329.7 ± 0.5	156 ± 7	−1.03 ± 0.13	59.6 ± 7.0	336.3 ± 0.6	177 ± 21	−1.16 ± 0.12	-2.19 ± 0.18
Lys157Ala[a]	34.8 ± 1.6	319.9 ± 0.8	109 ± 5	−0.06 ± 0.13	45.4 ± 3.3	325.6 ± 0.6	139 ± 10	0.59 ± 0.11	0.53 ± 0.17
Thr56Val[a]	52.8 ± 2.3	323.9 ± 0.7	163 ± 7	−0.46 ± 0.13	70.3 ± 4.5	331.5 ± 0.4	212 ± 14	−0.38 ± 0.08	−0.84 ± 0.15
Lys108Ala[a]	40.8 ± 1.3	322.0 ± 0.5	127 ± 4	−0.27 ± 0.11	50.9 ± 5.0	330.2 ± 0.7	154 ± 15	−0.16 ± 0.12	−0.43 ± 0.17
Asp100Asn[a]	48.5 ± 2.2	327.8 ± 0.8	148 ± 7	−0.85 ± 0.14	57.5 ± 6.8	332.7 ± 0.7	173 ± 20	−0.57 ± 0.13	−1.42 ± 0.19
Asp96Asn[a]	43.4 ± 1.9	328.4 ± 1.0	132 ± 6	−0.91 ± 0.16	52.8 ± 8.1	332.7 ± 1.1	159 ± 24	−0.57 ± 0.19	−1.48 ± 0.24

ΔH and T_m values are calculated by global fit of thermal denaturation curves followed by fluorescence, near-UV absorbance, and near and far-UV circular dichroism to a three-state model. Two curves of each technique have been included in the global fit. Errors in ΔH and T_m are provided by the program. ΔS and $\Delta\Delta G$ are calculated as described in Materials and Methods. Errors in ΔS and $\Delta\Delta G$ are calculated by propagation of the ΔH and T_m errors.

[a] These proteins bear a Trp120Phe mutation. The pWT protein (Trp120Phe) displays a stability and structure very similar to WT (not shown), and is the reference of this series.

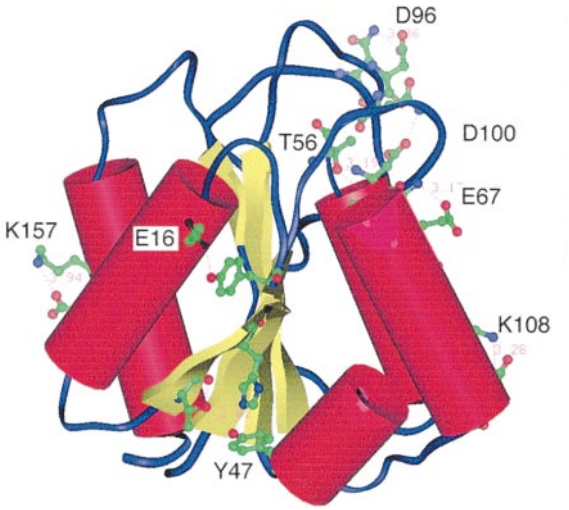

Figure 6. Ribbon diagram of the apoflavodoxin secondary structure showing the hydrogen bonds mutated. Labels indicate the mutated residues.

tein, and a global fit to the three-state mechanism has been performed for each mutant. The results of the global fits are shown in Table 2. For the native-to-intermediate equilibrium, changes in T_m (wild-type minus mutant) range from −4.4 to +4.6 K, and for the intermediate-to-unfolded equilibrium from −3.9 to +2.3 K (a detailed interpretation of the changes in T_m will have to await determination of the mutant X-ray structures).

As the relative effect of the mutations on the energetics of the two equilibria (native/intermediate and intermediate/unfolded) reports on the relative integrity of the mutated hydrogen bonds in the native and intermediate states, the structure of these two states can be compared at the residue level. Our approach is inspired by the φ-analysis, originally devised to obtain structural information on protein-folding transition states and folding kinetic intermediates,[42] and it is exemplified in Figure 7 for a mutation that decreases the stability of the protein. For each mutant, the difference in free energy of unfolding near the transition temperature (wild-type minus mutant) is calculated for the intermediate/unfolded equilibrium ($\Delta\Delta G_{IU}$) and for the overall native/unfolded equilibrium ($\Delta\Delta G_{NU}$; see Materials and Methods). We define the integrity of a given interaction in the intermediate (ϕ_I) as:

$$\phi_I = \Delta\Delta G_{IU}/\Delta\Delta G_{NU} \quad (1)$$

Any interaction, probed by mutation, that is similarly formed in the intermediate and in the folded state will give the same value for the two energy differences and, therefore, a ϕ_I value of 1. Conversely, any interaction that is present in the folded state but totally broken in the intermediate will give a $\Delta\Delta G_{IU}$ value of 0 and a ϕ_I value of 0. Finally, native interactions, partly formed in the intermediate, will give fractional integrity values. The assumptions of this method of analysis, which can be applied to mutations that stabilise or destabilise the protein, are similar to those described for the φ-analysis method.[42]

The ϕ_I values of the interactions probed by the mutations performed on apoflavodoxin hydrogen bonds are shown in Table 3. The integrity in the intermediate of two of the eight hydrogen bonds studied (Leu62/Glu67 and Asp74/Lys108) is difficult to assess because of the small $\Delta\Delta G$ values involved and will not be discussed further. Of the six remaining hydrogen bonds, whose mutation produced significant energy changes in the IU and NU equilibria allowing an accurate calculation of the integrity parameter, one is as formed in the intermediate as in the native state (Asp25/Lys157) and five are partly formed (40-60 %) in the intermediate, compared to the native state: the six hydrogen bonds are thus either partly or wholly

Table 3. Integrity, in the apoflavodoxin thermal intermediate, of mutated solvent exposed hydrogen bonds that connect secondary structure elements in the native state

Mutation	Hydrogen bond	H bond distance (Å)	Connected elements[a]	Integrity[b] (ϕ_I)
Glu16Gln	Tyr8/Glu16	2.74	β1-α1	0.53
Lys157Ala	Glu25/Lys157	3.27	α1-α5	1.11
Thr56Val	Thr56/Asp100	3.19	β3-α4	0.45
	Thr56/Ala101	2.98		
Glu67Ala	Leu62/Glu67	3.17	$Loop_{\beta3\alpha3}$-α3	[c]
Tyr47Phe	Tyr47/Thr32	3.02	β2-$loop_{\alpha2\beta3}$	0.63
	Tyr47/His34	2.68		
Lys108Ala	Asp74/Lys108	3.28	α3-α4	[c]
Asp100Asn	Asn97/Asp100	2.89	$Loop_{\beta4\alpha4}$-α4	0.40
Asp96Asn	Asp96/Asn128	3.06	$Loop_{\beta4\alpha4}$-$loop_{\beta5}$	0.39

[a] The subscripts of loops refer to the elements of secondary structure connected by the loops. $Loop_{\beta5}$ is a loop that splits the fifth β-strand in *Anabaena* flavodoxin.
[b] Defined as $\phi_I = \Delta\Delta G_{IU}/\Delta\Delta G_{NU}$, see Discussion.
[c] The integrity values corresponding to these hydrogen bonds are not reliable, due to the small energies involved in their calculation (see Table 2).

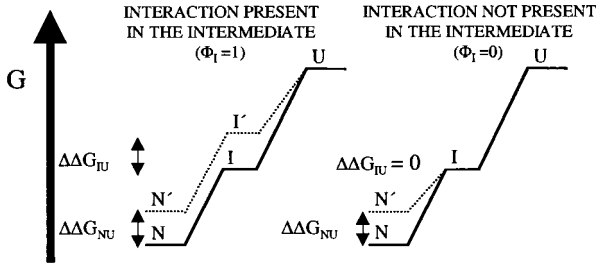

Figure 7. Energy diagrams showing different possible effects of mutations on the energetics of the two equilibria (NI and IU) and their relationship to the integrity parameter Φ_I.

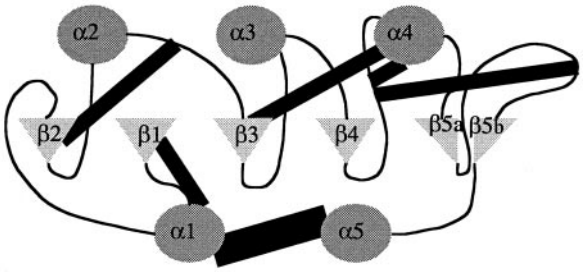

Figure 8. Apoflavodoxin secondary structure cartoon showing the integrity in the intermediate of the hydrogen bonds mutated. The thickest line connecting helices 1 and 5 represents a hydrogen bond fully formed in the intermediate. The other straight lines are partially formed hydrogen bonds.

formed in the intermediate. An alternative interpretation of these fractional ϕ values would be the presence of a mixture of numerous species with interactions formed to different degrees, but the high cooperativity shown by the apoflavodoxin thermal intermediate argues against it.

Intermediate native topology and native tertiary interactions

The hydrogen bonds probed by mutation are depicted in Figure 8. They connect neighbouring secondary-structure elements and act as inter-element contact probes. The integrity of the bonds in the intermediate, as probed by the energy analysis above, provides a low-resolution structure of this conformation. As hydrogen bonds are highly directional and short-range interactions, partial formation of a hydrogen bond is possible only when the residues involved are close and reasonably well oriented. Partly formed hydrogen bonds in the intermediate can thus be interpreted as indicating that the residues involved are similarly close and similarly oriented as in the native structure,[43] but the hydrogen bond is debilitated relative to the native state.

Flavodoxin is a three-layer α/β protein. One outer layer is formed by the N and C-terminal helices of the protein, helices 1 and 5 (Figure 8). The Asp25/Lys157 hydrogen bond (Table 3) connects the C-terminal region of helix 1 with the central region of helix 5. The integrity of the bond in the intermediate is of approximately 1. This indicates that the two helices are very probably formed (at least in part) and certainly in contact through those residues. This outer helical layer is, in turn, packed against the central sheet layer (Figure 8) as indicated by the integrity (0.53) of the Tyr8/Glu16 bond, which connects the N-terminal region of helix 1 with the centre of strand 1. Helix 1, thus, forms native-like contacts at its two ends (with helix 5 and with strand 1 respectively), indicating that most of the helix is formed in the thermal intermediate.

The central β-sheet layer is also connected to the other outer helical layer in flavodoxin, formed by helices 2, 3 and 4. One contact is established by strand 2, located at one edge of the sheet. This strand contains two residues (Thr32 and His34) hydrogen bonded, in native apoflavodoxin, to Tyr47, which is situated in a short loop connecting helix 2 and strand 3. The integrity of this double hydrogen bond in the intermediate (0.63) can be interpreted as arising from a simultaneous debilitation of the two hydrogen bonds or as indicating that one of the bonds is fully formed and the other one is broken. Whatever the case, strand 2 is connected to the loop. A second contact between the sheet layer and the helical layer of helices 2, 3 and 4, is established by Thr56 (at one end of strand 3), which is bonded to the main-chain nitrogen atoms of Asp100 and Ala101 (located at the N terminus of helix 4). As in the previous contact, the integrity of these bonds in the intermediate (0.45) may reflect a debilitation of both bonds, or that one remains as strong as in the native conformation and the other one is broken. Either possibility is an indication of debilitated native contacts between strand 3 and helix 4. A further hydrogen bond between the side-chain of Asp100 and that of Asn97 is debilitated but present in the intermediate, reflecting packing of helix 4 against the preceding loop connecting the helix to strand 4. The last hydrogen bond whose integrity in the thermal intermediate has been measured is established between Asp96 and Asn128, and connects two loops located in the interface between the central sheet layer and that of helices 2, 3 and 4. One of the loops (encompassing residues 118-140) is a long loop that splits the fifth strand and appears folded onto the rest of the protein in the native structure. The presence in the intermediate of a bond between one of the central residues of this loop (Asn128) and a residue located in a short loop

connecting two of the protein layers (Asp96) suggests that loop 118-140 may adopt in the intermediate a conformation similar to that shown in the native structure. On the whole, the data indicate that the thermal intermediate displays a close-to-native topology, with all the elements of secondary structure and loops that have been probed (three helices, three strands and three loops) properly oriented.

It should be mentioned that the hydrogen bonds analysed here (with the only exception of the Asn97/Asp100 bond) are tertiary interactions. As hydrogen bonds are more directional or shorter range than other interactions manifested in proteins, the evidence that all these surface-located hydrogen bonds connecting different regions in the protein are formed in the intermediate, although debilitated, suggests that the intermediate may contain a high percentage of native tertiary interactions. This is consistent with the substantial enthalpic stabilisation of the intermediate relative to the unfolded state that is revealed by the thermal denaturation experiments. It seems that, despite their clear spectroscopic differences, native proteins and molten globules may, in some cases, be closely related in both conformation and energetics, even at the residue level.

Conclusion

Protein molten globules[15] have spurred our imagination because, despite their fairly native secondary structure content, their spectroscopic properties (chiefly, lack of near-UV CD signal, and NMR poor signal dispersion) suggested that they were devoid of tertiary interactions. To our surprise, when their side-chain tertiary interactions have been studied on a residue basis, the emerging picture for both inner hydrophobic contacts[21-29] and surface hydrogen bonds (this work) is that molten globules may be quite close to native conformations. Supporting this view, molten globule energetics appears not very different from that of native proteins (reference 35 and this work). It seems thus that a quantitative reassessment of the structural meaning of the peculiar spectroscopic properties of molten globules is needed.

Materials and Methods

Mutagenesis, protein purification, protein quantification and apoprotein preparation

All *Anabaena* PCC 7119 flavodoxin variants (Glu16Ala, Tyr47Phe, The56Val, Glu67Ala, Asp96Asn, Asp100Asn, Lys108Ala and Lys157Ala) were prepared by the method of Deng & Nickoloff[44] directly on the expression plasmid pTrc 99a[31] and identified by sequencing the entire gene. The mutants Glu16Ala, Thr56Val, Asp96Asn, Asp100Asn, Lys108Ala, and Lys157Ala derive from pseudo-wild-type apoflavodoxin (pWT: Trp120Phe), that has been used as their reference. pWT apoflavodoxin is very similar in structure and stability to WT (not shown). All flavodoxin mutants were purified by an adaptation of the method of Fillat *et al.*[31] as described,[34] and the final preparations were homogeneous as shown by SDS-PAGE. To obtain the apoproteins, the flavin mononucleotide group was removed from the holoprotein by treatment with trichloroacetic acid.[45] Apoflavodoxin concentration was determined by recording A_{280nm} (extinction coefficients of 34.1 mM^{-1} cm^{-1} for the wild-type and mutants thereof,[34] and of 27.7 mM^{-1} cm^{-1} for the pseudo-wild-type and mutants thereof; unpublished data).

Thermal unfolding followed spectroscopically

Thermal unfolding curves were determined using four different spectroscopic techniques. For fluorescence, a ratio of emission (320/360 nm, with excitation at 280 nm) was used to minimise the strong baseline temperature-dependence. Apoprotein concentration was 2 μM and the buffer 50 mM Mops (pH 7). Circular dichroism and absorbance in the near-UV were simultaneously measured at 291 nm in a Jasco J-710 spectropolarimeter using 35-40 μM apoflavodoxin in 50 mM Mops (pH 7.0). Circular dichroism in the far-UV (222 nm) was recorded with 1 μM apoprotein in 5 mM Mops (pH 7.0) (plus NaCl to reach the ionic strength used in the other techniques). Data of thermal denaturation were preliminarily fitted to the equation for a two-state equilibrium:

$$S = (S_{N0} + m_N T + (S_{U0} + m_U T)\exp(-\Delta G(T)/RT))/ (1 + \exp(-\Delta G(T)/RT)) \quad (2)$$

where S_{N0} and S_{U0} represent the spectroscopic signals of the native and unfolded states at 0 K, and m_F and m_U are the temperature dependence of those signals. The Gibbs energy function, $\Delta G(T)$, follows equation (3):

$$\Delta G(T) = \Delta H(T_m) \times (1 - T/T_m) - \Delta C_p \times ((T_m - T) + T\ln(T/T_m)) \quad (3)$$

with ΔC_p and $\Delta H(T_m)$ being the folding heat capacity and enthalpy changes, respectively. Fitting spectroscopic unfolding curves to equation (1) provides accurate T_m values, reasonable ΔH values and, usually unreliable, ΔC_p values.[46]

Thermal unfolding followed by calorimetry

DSC experiments were performed with a VP-DSC calorimeter from MicroCal (Northhampton, MA). Apoflavodoxin solutions were prepared by exhaustive dialysis against buffer. The samples were degassed at room temperature before the calorimetric experiments. Calorimetric cells (operating volume ~0.5 ml) were kept under an excess pressure of 30 psi (1 psi ≈ 6.9 kPa) to prevent degassing during the scan. Experiments were carried out at a scan-rate of 0.5 K min^{-1} or 1.5 K min^{-1}. In all measurements, the buffer from the last dialysis step was used in the reference of the calorimeter. Several buffer-buffer baselines were obtained before each run with a protein solution in order to ascertain proper equilibration of the calorimeter, and an additional buffer-buffer baseline was obtained after each protein run to check that no significant change in instrumental baseline had occurred. In selected experiments, a reheating run was carried out to determine the reversibility of the denaturation process (the reversibility was about 90 % when the first scan was

stopped immediately after the transition). Fittings of the theoretical models to the heat capacity profiles were performed using programs written by us in the MLAB environment (Civilized Software, Inc.). All fittings were done to the absolute heat capacity values, which were calculated using programs provided by MicroCal. The general approach used in the two-state fittings is as described,[47] assuming the unfolding enthalpy value can be taken as a constant within the narrow temperature range of the transition (and assigned to the transition temperature); on the other hand, the pre and post-transition baselines span a wider temperature range and, therefore, the temperature-dependence of the heat capacities for the native and unfolded states was specifically taken into account using expressions linear in temperature. Three-state fittings were carried out in a similar manner but using the three-state expressions derived from the well-known, multi-state, partition-function-based approach (see, for instance, reference 39). The heat capacity of the intermediate state was assumed to be related to the heat capacities of the native and unfolded states through:

$$C_p(\mathrm{I}) = (1 - Z) \times C_p(\mathrm{N}) + Z \times C_p(\mathrm{U}) \quad (4)$$

where Z describes the similarity of the heat capacity of the intermediate state to those of the native and unfolded states ($Z = 1$ if $C_p(\mathrm{I}) = C_p(\mathrm{U})$ and $Z = 0$ if $C_p(\mathrm{I}) = C_p(\mathrm{N})$). This calorimetric Z value was fixed in every specific fitting. We obtained visually excellent fittings using Z values between 0 and 1 although, for $Z > 0.6$, the calculated temperature-dependencies of $C_p(\mathrm{N})$ and $C_p(\mathrm{U})$ crossed within the studied temperature range.

Three-state global analysis of the spectroscopic thermal unfolding curves

Global fitting of the four thermal unfolding curves of each protein (fluorescence, far-UV CD, near-UV CD, and near-UV absorbance) to a three-state model involving native (N), intermediate (I) and unfolded (U) conformations was performed with the program MLAB (from Civilized Software).

The optical data were converted[41] to apparent fractions of unfolded protein, F_{app}, using equation (5):

$$F_{app} = (Y - Y_N)/(Y_U - Y_N) \quad (5)$$

where Y is the observed signal, and Y_U and Y_N are the optical signals for unfolded and native protein, respectively. Both Y_N and Y_U were observed to depend linearly on the temperature, and the linear dependence was assumed to hold in the transition region. F_{app} can be recast in terms of the fractional populations of the intermediate and unfolded states (F_I and F_U) as:

$$F_{app} = F_U + ZF_I \quad (6)$$

where $Z = (Y_I - Y_N)/(Y_U - Y_N)$. The Z parameter describes the degree to which the I state spectroscopically resembles the U state.

Using the relationship between fractional populations and equilibrium constants, and the balance of fractions ($F_U + F_I + F_U = 1$), F_{app} can be calculated from equation (7):

$$F_{app} = \frac{[Z + \exp(-\Delta G_{IU}(T)/RT)]}{[1 + \exp(\Delta G_{NI}(T)/RT) + \exp(-\Delta G_{IU}(T)/RT)]} \quad (7)$$

where:

$$\Delta G_{NI}(T) = \Delta H_{NI}(T_{mNI}) \times [1 - T/T_{mNI}] - \Delta C_{pNI}[(T_{mNI} - T) + T \ln(T/T_{mNI})] \quad (8)$$

$$\Delta G_{IU}(T) = \Delta H_{IU}(T_{mIU}) \times [1 - T/T_{mIU}] - \Delta C_{pIU}[(T_{mIU} - T) + T \ln(T/T_{mIU})] \quad (9)$$

In each fitting, all thermodynamic parameters were globally constrained, while the optical properties of the intermediate (the Z values) were allowed to vary among the different spectroscopic techniques.

Analysis of the structure of the apoflavodoxin thermal intermediate

The free energy difference upon mutation (wild-type minus mutant) for the NI and IU transitions were calculated using equations (10) and (11):[41]

$$\Delta\Delta G_{NI} = \Delta S_{mNI}(\mathrm{WT}) \times \Delta T_{mNI}(\mathrm{WT - mutant}) \quad (10)$$

$$\Delta\Delta G_{IU} = \Delta S_{mIU}(\mathrm{WT}) \times \Delta T_{mIU}(\mathrm{WT - mutant}) \quad (11)$$

The degree of formation in the intermediate of a probed interaction (relative to the native state) was calculated from equation (1) (see Results and Discussion) where $\Delta\Delta G_{NU} = \Delta\Delta G_{NI} + \Delta\Delta G_{IU}$.

Acknowledgements

We acknowledge financial support from grants PB97-1027 and PB96-1439 (DGES, Spain), and P15/97 (CONSI + D, DGA, Spain). M.P.I. has been supported by an FPU fellowship (Spain). M.G-M. is a recipient of a predoctoral fellowship from the Junta de Andalucia.

References

1. Fersht, A. R. & Serrano, L. (1993). Principles of protein stability derived from protein engineering experiments. *Curr. Opin. Struct. Biol.* **3**, 75-83.
2. Shaw, A. & Bott, R. (1996). Engineering enzymes for stability. *Curr. Opin. Struct. Biol.* **6**, 546-550.
3. Pace, C. N., Shirley, B. A., McNutt, M. & Gajiwala, K. (1996). Forces contributing to the conformational stability of proteins. *FASEB J.* **10**, 75-83.
4. Richards, F. M. (1997). Protein stability: still an unsolved problem. *Cell Mol. Life Sci.* **53**, 790-802.
5. Honig, B. (1999). Protein folding: from the Levinthal paradox to structure prediction. *J. Mol. Biol.* **293**, 283-293.
6. Nicholson, H., Becktel, W. J. & Matthews, B. W. (1988). Enhanced protein thermostability from designed mutations that interact with α-helix dipoles. *Nature*, **336**, 651-656.
7. Matsumura, M., Signor, G. & Matthews, B. W. (1989). Substantial increase of protein stability by multiple disulphide bonds. *Nature*, **342**, 291-293.
8. Zhang, X-J., Baase, W. A., Shoichet, B. K., Wilson, K. P. & Matthews, B. W. (1995). Enhancement of protein stability by the combination of point

mutations in T4 lysozyme is additive. *Protein Eng.* **8**, 1017-1022.
9. Shih, P. & Kisch, J. F. (1995). Design and structural analysis of an engineered thermostable chicken lysozyme. *Protein Sci.* **4**, 2063-2072.
10. Villegas, V., Viguera, A. R., Aviles, F. X. & Serrano, L. (1996). Stabilization of proteins by rational design of alpha-helix stability using helix/coil transition theory. *Fold. Des.* **1**, 29-34.
11. Williams, J. C., Zeelen, J. P., Neubauer, G., Vriend, G., Backmann, J., Michels, P. A. M., Lambeir, A.-M. & Wierenga, R. K. (1999). Structural and mutagenesis studies of leishmania triosephosphate isomerase: a point mutation can convert a mesophilic enzyme into a superstable enzyme without losing catalytic power. *Protein Eng.* **12**, 243-250.
12. Grimsley, G. R., Shaw, K. L., Fee, L. R., Alston, R. W., Huyghues-Despointes, B. M., Thurlkill, R. L., Scholtz, J. M. & Pace, C. N. (1999). Increasing protein stability by altering long-range Coulombic interactions. *Protein Sci.* **8**, 1843-1849.
13. Loladze, V. V., Ibarra-Molero, B., Sanchez-Ruiz, J. M. & Makhatadze, G. I. (1999). Engineering a thermostable protein via optimisation of charge-charge interactions in the protein surface. *Biochemistry,* **38**, 16419-16423.
14. Roder, H. & Colon, W. (1997). Kinetic role of early intermediates in protein folding. *Curr. Opin. Struct. Biol.* **7**, 15-28.
15. Ptitsyn, O. B. (1992). The molten globule state. In *Protein Folding* (Creighton, T. E., ed.), pp. 243-300, Freeman, New York.
16. Wolynes, P. G., Onuchic, J. N. & Thirumalai, D. (1995). Navigating the folding routes. *Science,* **267**, 1619-1620.
17. Jennings, P. A. & Wright, P. E. (1993). Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin. *Science,* **262**, 892-896.
18. Balbach, J., Forge, V., van Nuland, N. A. J., Winder, S. L., Hore, P. J. & Dobson, C. M. (1995). Following protein folding in real time using NMR spectroscopy. *Nature Struct. Biol.* **2**, 866-870.
19. Raschke, T. M. & 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.
20. Guijarro, J. I., Sunde, M., Jones, J. A., Campbell, I. D. & Dobson, C. M. (1998). Amyloid fibril formation by an SH3 domain. *Proc. Natl Acad. Sci. USA,* **95**, 4224-4228.
21. Wu, L. C. & Kim, P. S. (1998). A specific hydrophobic core in the alpha-lactalbumin molten globule. *J. Mol. Biol.* **280**, 175-182.
22. Kiefhaber, T. & Baldwin, R. L. (1995). Intrinsic stability of individual alpha helices modulates structure and stability of the apomyoglobin molten globule form. *J. Mol. Biol.* **252**, 122-132.
23. Uchiyama, H., Perez-Prat, E. M., Watanabe, K., Kumagai, I. & Kuwajima, K. (1995). Effects of amino acid substitutions in the hydrophobic core of alpha-lactalbumin on the stability of the molten globule state. *Protein Eng.* **8**, 1153-1161.
24. Kay, M. S., Ramos, C. H. I. & Baldwin, R. L. (1999). Specificity of native-like interhelical hydrophobic contacts in the apomyoglobin intermediate. *Proc. Natl Acad. Sci. USA,* **96**, 2007-2012.
25. Song, J., Bai, P., Luo, L. & Peng, Z. Y. (1998). Contribution of individual residues to formation of the native-like tertiary topology in the alpha-lactalbumin molten globule. *J. Mol. Biol.* **280**, 167-174.
26. Marmorino, J. L., Lehti, M. & Pielak, G. J. (1998). Native tertiary structure in an A-state. *J. Mol. Biol.* **275**, 379-388.
27. Schulman, B. A. & Kim, P. S. (1996). Proline scanning mutagenesis of a molten globule reveals non-cooperative formation of protein's overall topology. *Nature Struct. Biol.* **3**, 682-687.
28. Kay, M. S. & Baldwin, R. L. (1996). Packing interactions in the apomyoglobin folding intermediate. *Nature Struct. Biol.* **3**, 429-445.
29. Luo, Y., Kay, M. S. & Baldwin, R. L. (1997). Cooperativity of folding of the apomyoglobin pH 4 intermediate studied by glycine and proline mutations. *Nature Struct. Biol.* **4**, 925-930.
30. Matouschek, A., Serrano, L. & Fersht, A. R. (1992). The folding of an enzyme. IV. Structure of an intermediate in the refolding of barnase analysed by a protein engineering procedure. *J. Mol. Biol.* **224**, 819-835.
31. Fillat, M. F., Borrias, W. E. & Weisbeerk, P. J. (1991). Isolation and overexpression in *Escherichia coli* of the flavodoxin gene from *Anabæna* PCC 7119. *Biochem. J.* **280**, 187-191.
32. Genzor, C. G., Perales-Alcón, A., Sancho, J. & Romero, A. (1996a). Closure of a tyrosine/tryptophan aromatic gate leads to a compact fold in apoflavodoxin. *Nature Struct. Biol.* **3**, 329-332.
33. Maldonado, S., Lostao, A., Irún, M. P., Fernández-Recio, J., Genzor, C. G., González, E. B., Rubio, J. A., Luquita, A., Daoudi, F. & Sancho, J. (1998b). Apoflavodoxin: structure, stability, and FMN binding. *Biochimie,* **80**, 813-820.
34. Genzor, C. G., Beldarraín, A., Gomez-Moreno, C., Lopez-Lacomba, J. L., Cortijo, M. & Sancho, J. (1996b). Conformational stability of apoflavodoxin. *Protein Sci.* **5**, 1376-1388.
35. Maldonado, S., Jiménez, M. A., Langdon, G. M. & Sancho, J. (1998a). Cooperative stabilisation of a molten globule apoflavodoxin fragment. *Biochemistry,* **37**, 10589-10596.
36. Fernández-Recio, J., Romero, A. & Sancho, J. (1999). Energetics of a hydrogen bond (charged and neutral) and of a cation-π interaction in apoflavodoxin. *J. Mol. Biol.* **290**, 319-331.
37. Lostao, A., El Harrous, M., Daoudi, F., Romero, A., Parody-Morreale, A. & Sancho, J. (2000). Dissecting the energetics of the apoflavodoxin-FMN complex. *J. Biol. Chem.* **275**, 9518-9526.
38. van Mierlo, C. P., van Dongen, W. M., Vergeldt, F., van Berkel, W. J. & Steensma, E. (1998). The equilibrium unfolding of *Azotobacter vinelandii* apoflavodoxin II occurs *via* a relatively stable folding intermediate. *Protein Sci.* **7**, 2331-2344.
39. Sanchez-Ruiz, J. M. (1995). Differential scanning calorimetry of proteins. In *Subcellular Biochemistry* (Biswas, B. B. & Siddharta, R., eds), vol. 24, pp. 133-176, Plenum Press, New York.
40. Plotnikov, V. V., Brandts, J. M., Lin, L.-N. & Brandts, J. F. (1997). A new ultrasensitive scanning calorimeter. *Anal. Biochem.* **250**, 237-244.
41. Luo, J., Iwakura, M. & Matthews, C. R. (1995). Detection of a stable intermediate in the thermal unfolding of a cysteine-free form of dihydrofolate reductase from *Escherichia coli*. *Biochemistry,* **34**, 10669-10675.

42. Fersht, A. R., Matouscheck, A. & Serrano, L. (1992). The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding. *J. Mol. Biol.* **224**, 771-782.
43. Sancho, J., Meiering, E. M. & Fersht, A. R. (1991). Mapping transition states of protein unfolding by protein engineering of ligand-binding sites. *J. Mol. Biol.* **221**, 1007-1014.
44. Deng, W. P. & Nickoloff, J. A. (1992). Site-directed mutagenesis of virtually any plasmid by eliminating a unique site. *Anal. Biochem.* **200**, 81-88.
45. Edmondson, D. E. & Tollin, G. (1971). Chemical and physical characterisation of the Shethna flavoprotein and kinetics and thermodynamics of flavin analogue binding to the apoprotein. *Biochemistry*, **10**, 124-132.
46. Pace, C. N., Shirley, B. A. & Thomson, J. A. (1989). Measuring the conformational stability of a protein. In *Protein Structure: A Practical Approach*, pp. 311-330, IRL Press, Oxford.
47. Ibarra-Molero, B., Makhatadze, G. I. & Sanchez-Ruiz, J. M. (1999). Cold denaturation of ubiquitin. *Biochim. Biophys. Acta,* **1429**, 384-390.

Edited by A. R. Fersht

(Received 29 September 2000; received in revised form 2 January 2001; accepted 4 January 2001)