doi:10.1016/j.jmb.2004.08.081 J. Mol. Biol. (2004) 344, 239–255

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Structure of Stable Protein Folding Intermediates by Equilibrium φ-Analysis: The Apoflavodoxin Thermal Intermediate

Luis A. Campos[1], Marta Bueno[1], Jon Lopez-Llano[1] María Ángeles Jiménez[2] and Javier Sancho[1]*

[1]*Biocomputation and Complex Systems Physics Institute and Department Bioquímica y Biología Molecular y Celular Fac. Ciencias, University Zaragoza, 50009 Zaragoza Spain*

[2]*Instituto de Química-Física Rocasolano, CSIC, Serrano-119 28006 Madrid, Spain*

**Corresponding author*

Protein intermediates in equilibrium with native states may play important roles in protein dynamics but, in cases, can initiate harmful aggregation events. Investigating equilibrium protein intermediates is thus important for understanding protein behaviour (useful or pernicious) but it is hampered by difficulties in gathering structural information. We show here that the φ-analysis techniques developed to investigate transition states of protein folding can be extended to determine low-resolution three-dimensional structures of protein equilibrium intermediates. The analysis proposed is based solely on equilibrium data and is illustrated by determination of the structure of the apoflavodoxin thermal unfolding intermediate. In this conformation, a large part of the protein remains close to natively folded, but a 40 residue region is clearly unfolded. This structure is fully consistent with the NMR data gathered on an apoflavodoxin mutant designed specifically to stabilise the intermediate. The structure shows that the folded region of the intermediate is much larger than the proton slow-exchange core at 25 °C. It also reveals that the unfolded region is made of elements whose packing surface is more polar than average. In addition, it constitutes a useful guide to rationally stabilise the native state relative to the intermediate state, a far from trivial task.

Keywords: protein stability; protein folding; equilibrium intermediate; thermal unfolding; φ-analysis

Introduction

Proteins are not all simple two-state systems. Many adopt, under certain conditions, stable folded conformations that are different from the functionally relevant native state. These conformations, usually known as equilibrium intermediates, may appear under unusual external conditions (i.e. non-physiological pH, pressure or temperature), in the presence of high concentrations of certain molecules (denaturants, salts), or as a consequence of mutations.[1–9] The occurrence of equilibrium intermediates is sometimes associated with well-characterized stress phenomena, where specific physiological responses tend to reduce their harmful effects[10–12] and, in some cases, they are linked to human diseases, such as spongiform encephalopathy and other types of amyloidosis.[13,14] For some proteins, however, physiological roles have been postulated for their intermediate conformations.[8,15–18] It is very clear, at present, that protein equilibrium intermediates must be actively investigated to achieve a proper understanding of protein behaviour, to be able to restore, when required, the native functions that are lost in these conformations, and to combat the harmful actions that some intermediates seem to exert *in vivo*.

One major problem with respect to achieving those goals is that the structural study of intermediates is much more difficult than that of native conformations. So far, X-ray crystallography has been of little use because the intrinsic greater flexibility associated with the intermediates precludes their crystallization. NMR has thus been the technique of choice[19–24] and yet obtaining high-resolution structures of intermediates has proven difficult due to

Abbreviations used: SAS, solvent-accessible surfaces; CD, circular dichroism.

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

extensive signal overlapping. In addition, NMR studies are further complicated when the percentage of intermediate under conditions of maximal accumulation is low compared to those of the native and/or unfolded states. As for the lower-resolution techniques, such as circular dichroism (CD), Fourier transform infrared (FTIR), fluorescence resonance energy transfer (FRET), light-scattering, etc., although useful, they can hardly provide enough structural information at the residue level. There is thus a need for new techniques to study protein equilibrium intermediates with reasonable detail, certainly at a residue level, and able to investigate the structure of those intermediates that never constitute the major species at equilibrium. We show here that φ-analysis,[25] originally devised to investigate transition states of protein unfolding and transient folding intermediates[26,27] can be adapted to obtain low-resolution structures of equilibrium intermediates. This is so because the essence of φ-analysis (that the unknown structure of a conformation at equilibrium with two additional conformations of known structure can be inferred, using wild-type and mutant proteins, from the analysis of the energetics of the two equilibria) perfectly applies to protein equilibrium intermediates.[28] In the φ-analysis of equilibrium intermediates, no kinetic data are required and all the information derives from global analysis of the data obtained by monitoring the unfolding equilibrium with different techniques (spectroscopic or otherwise).

We demonstrate here the feasibility of the method by solving, at low resolution, the structure of the intermediate that accumulates in the thermal unfolding of the apoflavodoxin from *Anabaena* PCC 7119. A comparison of this structure with that of the native state[29] and with the hydrogen slow-exchange core of the protein[30] allows us to draw a clear picture of the regions of the intermediate that remain close to native, and gives insight into the potential relationship between thermal intermediates and slow exchange cores. In addition, the structure of the apoflavodoxin intermediate offers clear clues that permit the rational, specific stabilisation of the native conformation relative to the intermediate,[31] not just relative to the unfolded state.

Results

Structural integrity of mutants

Wild-type and 31 apoflavodoxin mutants have been analysed to delineate the structure of the thermal equilibrium intermediate. Eighteen mutations have been newly made that are distributed over the flavodoxin structure (Figure 1). The mutations (L6A, I22V, V31A, D43A, I51V, I52V, S71A, A84G, N97A, Q99A, I104V, S110A, V117A, T122S, V139A, L143A, I156V and V160A) are as conservative as possible and shorten the side-chain without changing the branching. The interactions

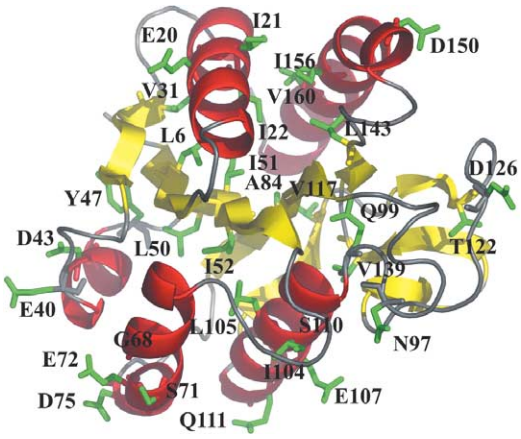

Figure 1. Ribbon representation of the apoflavodoxin structure (PDB code 1ftg) showing in green the residues that have been mutated in order to determine a low-resolution structure of the equilibrium intermediate of its thermal unfolding. Of the 31 mutations analysed, 29 gave rise to equilibrium φ-values of between 0 and 1.

that are broken by the mutations are shown in Table 1. The integrity of these mutants has been analysed by near-UV and far-UV CD (Figure 2). The far-UV CD spectra of the mutants are in most cases quite close to that of wild-type. For I22V, V31A, I51V, I52V, S71A, A84G, Q99A, S110A and L143A the spectrum is virtually identical; L6A, D43A, N97A, I104V, V117A, T122S, V139A and V160A show a slightly more negative signal at 209 nm; while the signal at 222 nm is identical with wild-type. Since there seem to be aromatic contributions[32] to the far-UV CD spectrum of apoflavodoxin,[29] these minute differences should not be taken as an indication of changes in secondary structure. The spectrum of I156V, however, displays a moderate loss of signal at 222 nm, which is consistent with the fact that, due to its low level of stability, some 8% of the molecules are already unfolded in 0 M urea (Table 2). Figure 2(b) shows for comparison the far-UV CD spectrum of the denatured wild-type protein at 80 °C. The spectrum suggests that thermal unfolding, as found for other proteins, might not disrupt the secondary structure as completely as urea-unfolding does.[33] We notice, nevertheless, that the changes in ΔG_{ND} introduced by the different mutations analysed are essentially the same when they are calculated from analysis of the thermal denaturation curves or from the urea-denaturation curves (see below). This suggests that, at least for the residues mutated, the thermally unfolded state of apoflavodoxin is thermodynamically equivalent to the chemically unfolded state (Figure 3).

Changes in the near-UV CD spectrum upon mutation are usually expected to be more common because the contributions to this spectral region arise from subtle interactions in the environment of aromatic residues. Most of the mutants analysed show spectra very similar to wild-type (in a few, L6A, N97A, I156V and V160A, the shape is identical

Table 1. Structural description of mutations implemented in apoflavodoxin

Mutation	Position	H-bond partners	Van der Waals contacts (<4 Å)	Structure elements involved
L6A	Strand 1		L33/I51/F86	Strand 2/strand 3/strand 4
I22V	Helix 1		V18/F86	Helix 1/strand 4
V31A	Strand 2		R23/F26/L33	Helix 1/strand 2
D43A	Helix 2	N (E40)	A39/E40	Loop
I51V	Strand 3		I4/F26	Strand 1/helix 1
I52V	Strand 3		W66	Helix 3
S71A	Helix 3	O (E67)/O (G68)		Helix 3
A84G	Strand 4		L50/V117/W159/L163	Strand 3/strand 5a/helix 5
N97A	Loop	OD2 (D100)	I59/D100	Loop/helix 4
Q99A	Loop	O (G87)/O (L143)	G87/T88/G89/I102/A142/ L143	Strand 4/helix 4/strand 5b
I104V	Helix 4		D100	Helix 4
S110A	Helix 4	O (E106)/O (E107)		Helix 4
V117A	Strand 5a		V83/L163/E166/F167	Strand 4/helix 5
T122S	Strand 6	O (G136)	L133/G136/K137/F138	Strand 7/strand 8
V139A	Strand 8		I102/E106/T116/Y119/R134	Helix 4/strand 5a/strand 6/strand 7
L143A	Strand 5b		L141	Strand 5b
I156V	Helix 5		L141/A142/L143/T152	Strand 5b/helix 5
V160A	Helix 5		E25/F26/I156/K164	Helix 1/helix 5
Y47F[a]	Loop	OG1 (T32)/NE2 (H34)		Strand 2
L50A[b]	Strand 3		G5/F7/L44/Y47/K81	Strand 1/helix 2/loop/strand 4
L105A[b]	Helix 4		I52/C54/L62/Y70/A101	Strand 3/loop/helix 3/helix 4
I21G[c]	Helix 1		S17/V18	Helix 1
G68A[c]	Helix 3		S64/D65	Helix 3
E107G[c]	Helix 4		G103/I104	Helix 4
Q111A[c]	Helix 4	NZ (K108)		Helix 4

In addition, the following mutants have been analysed:[31] E20K, E40K, E72K, D75K, D126K and D150K.
[a] Previously analysed.[28]
[b] Mutants (M.B. & J.S., unpublished results).
[c] Mutants (J.L.-L & J.S., unpublished results).

but the intensity is a bit lower). The CD data (both in the far and near-UV regions) thus indicate that, with the possible exception of I156V, all the other mutants display a structure similar to wild-type. In addition to these mutations, data corresponding to the following mutants, already reported[28] (Y47F) or related to work in progress in the lab (L50A, L105A, I21G, G68A, E107G and Q111A: MB, J.L.-L. & J.S., unpublished results) have been used to delineate the structure of the intermediate. They all display CD spectra quite close to those of wild-type (not shown). Besides, data corresponding to the E20K, E40K, E72K, D75K, D126K and D150K mutants have also been used, as will be explained. These mutants are all well folded.[31]

Changes in global stability: urea-denaturation

Although the thermal unfolding of apoflavodoxin is three-state,[28] its equilibrium urea-denaturation follows a simple two-state mechanism[29,30] and thus the overall free energy difference between the folded and unfolded state can be determined from urea-denaturation curves (Figure 4). The overall stability of the mutants, as calculated from fittings to equation (4) (see Methods) of the curves shown in Figure 4, is reported in Table 2. As the expected outcome of mutations that have been designed to break specific interactions in the folded state, most mutant proteins are less stable than wild-type. Two mutants, nevertheless, are more stable. One is D43A, whose small stabilisation might be due to some relief of electrostatic repulsions within the protein.[31] The larger stabilisation observed in the other mutant, Q99A, is however unexpected because, as the side-chain is buried, the mutation should lead to a loss in van der Waals contacts and in hydrophobic effect. As a possible clue, inspection of the structure indicates that while the glutamine side-chain NH is hydrogen bonded to main-chain groups in the protein, the side-chain CO is only bonded to a buried water molecule. The greater stability of the Q99A mutant could be explained if the desolvation of the side-chain CO upon folding is not properly compensated by its bonding to the buried water molecule. In that case, the replacement of glutamine by alanine would be energetically favourable.

Changes in the stability of the native state relative to the thermal equilibrium intermediate and in the stability of the intermediate relative to the denatured state: three-state thermal unfolding

We have established that no aggregation occurs during the thermal unfolding of wild-type apoflavodoxin by recording thermal unfolding curves using 2 μM and 20 μM protein solutions. The curves yield the same thermodynamic parameters (not shown), and the mutants are assumed to behave similarly. The reversibility of apoflavodoxin thermal unfolding was carefully investigated in a previous work.[28] In differential scanning

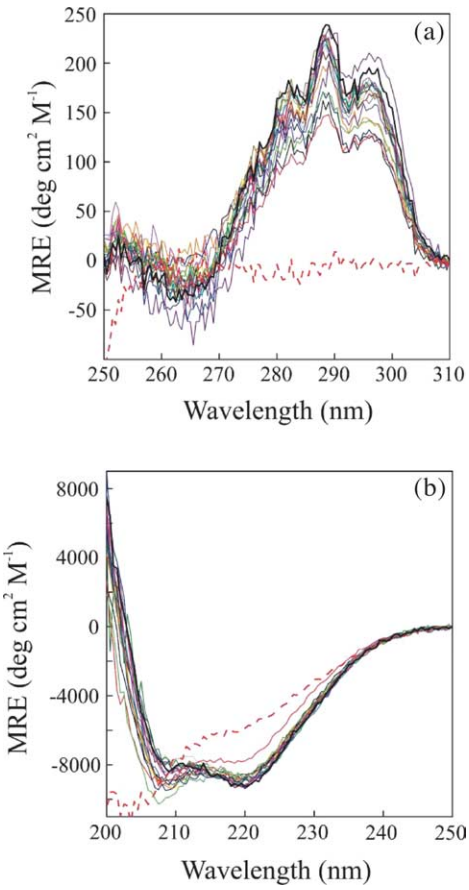

Figure 2. (a) Near-UV and (b) far-UV CD spectra of wild-type (black thick line), L6A (black line), I22V (light grey line), V31A (dark grey line), D43A (yellow line), I51V (cyan line), I52V (magenta line), S71A (blue line), A84G (dark yellow line), N97A (dark blue line), Q99A (purple line), I104V (wine line), S110A (olive line), V117A (dark cyan line), T122S (orange line), V139A (violet line), L143A (pink line), I156V (red line) and V160A (green line) proteins. The spectrum of the denatured wild-type protein at 80.0 °C (red thick broken line) is also shown for comparison. All other spectra were recorded in 50 mM Mops (pH 7.0), at 25.0 °C.

calorimetry (DSC) experiments, 90% of the unfolded protein was successfully renatured upon cooling and, in spectroscopically followed thermal unfolding (performed at lower concentrations) the signal recovery was even larger when the unfolding was stopped immediately after the transition. The same behaviour was observed for eight additional mutants that were then analysed. As for the mutants described here, we have typically recovered between 85% and 97% of the native near-UV ellipticity (which is fully lost upon heating in all cases). Their thermal unfolding is thus also reversible. For each protein variant, four thermal denaturation curves monitored by different spectroscopic techniques have been analysed globally to calculate the effect of the mutations on the energetics of the N-to-I and of the I-to-D equilibria. The data for each protein are globally fitted to equation (6), as explained in Methods (Figure 4 shows the fit for the wild-type protein). A problem encountered commonly in the analysis of thermal unfolding curves followed spectroscopically is that the ΔC_p values obtained are not usually reliable.[34] Although this hardly lowers the accuracy of the calculated $\Delta H(T_m)$ and T_m values, it complicates the extrapolation of the data beyond the transition region. In order to calculate equilibrium ϕ-values for the different mutations at a common reference temperature (see below) we need to make short extrapolations and therefore special care has been taken to improve the quality of the ΔC_p fitted values. To that end, the ΔC_p values have been constrained in the fitting to be within the standard deviation interval around the mean of the ΔC_p values calculated without restrictions for 40 different apoflavodoxin mutants studied either in this or in other related works (not shown). Thus, ΔC_{p1} is bound to be from 5.4 kJ K^{-1} mol^{-1} to 7.6 kJ K^{-1} mol^{-1} and ΔC_{p2} from 0.0 kJ K^{-1} mol^{-1} to 2.6 kJ K^{-1} mol^{-1} in the fittings of the wild-type and mutant proteins, which does not significantly change the fitted ΔH and T_m values obtained (not shown). Similar to what is observed in urea denaturation, most mutations destabilise the protein and lower the temperatures for the two transitions (Table 3). Some mutations, nevertheless, increase the melting temperature of the first transition (D43A, Q99A and T122S), and/or of the second one (D43A, S71A, Q99A and T122S). The changes in free energy differences of the N-to-I and I-to-D equilibria caused by the mutations can be, in principle, calculated at any temperature using equations (7) and (8). However, long extrapolations from the transition temperatures should be avoided and, if possible, simplified equations should be used to minimise the uncertainty introduced by the less accurate ΔC_p values.

In a previous work,[28] the changes caused by the mutations in the free energies of the two equilibria were calculated, respectively, in the vicinity of the corresponding melting temperatures, where the second terms in equations (7) and (8) are insignificant with respect to the first terms and the equations can be simplified to:

$$\Delta G = \Delta H(1 - T/T_m) \quad (1)$$

Changes in stabilities at the melting temperatures of the wild-type protein were then calculated, using equation (2), by subtracting wild-type ΔG_1 or ΔG_2 values from those of a given mutant:

$$\Delta\Delta G = \Delta H(1 - T_m^{WT}/T_m^{mut}) \quad (2)$$

where $T_m^{WT(mut)}$ is the melting temperature of the wild-type (mutant) protein in either the N-to-I (T_{m1}) or the I-to-D (T_{m2}) transition and ΔH is the enthalpy change of the mutant at its T_m. In that way, $\Delta\Delta G_1$ and $\Delta\Delta G_2$ were calculated, respectively, at the T_{m1} and T_{m2} of wild-type.

One obvious limitation of this approach is that the stability changes are calculated at two different temperatures, which complicates the physical interpretation of the ϕ-values that are derived

Table 2. Urea-denaturation of wild-type and mutant apoflavodoxins

Protein	$U_{1/2}$[a] (M)	m[a] (kcal mol^{-1} M^{-1})	$\Delta\Delta G_{fold}$[b] (kcal mol^{-1})	$\Delta\Delta G_{fold}$[c] (kcal mol^{-1})
Wt	2.067±0.004	2.44±0.03	–	
L6A	0.777±0.010	1.96±0.06	+3.53±0.09	+2.94±0.06
I22V	1.459±0.009	2.13±0.05	+1.94±0.11	+1.39±0.04
V31A	1.314±0.009	2.02±0.05	+2.39±0.10	+1.72±0.04
D43A	2.109±0.010	2.43±0.08	−0.09±0.18	−0.10±0.02
I51V	1.267±0.013	2.42±0.11	+1.98±0.16	+1.83±0.05
I52V	1.453±0.016	2.52±0.13	+1.38±0.21	+1.40±0.05
S71A	1.765±0.026	2.49±0.18	+0.65±0.33	+0.69±0.06
A84G	1.279±0.009	2.25±0.07	+2.17±0.11	+1.80±0.04
N97A	1.864±0.007	2.27±0.05	+0.83±0.11	+0.46±0.02
Q99A	2.333±0.008	2.90±0.09	−1.72±0.22	−1.72±0.22[d]
I104V	1.822±0.005	2.41±0.04	+0.67±0.10	+0.56±0.02
S110A	1.788±0.005	2.25±0.03	+1.03±0.09	+0.64±0.02
V117A	1.304±0.005	2.30±0.03	+2.05±0.08	+1.74±0.04
T122S	1.933±0.008	2.42±0.06	+0.37±0.14	+0.31±0.02
V139A	1.684±0.017	2.10±0.09	+1.51±0.17	+0.87±0.04
L143A	2.021±0.014	2.43±0.12	+0.14±0.25	+0.10±0.03
I156V	0.987±0.017	1.47±0.06	+3.60±0.10	+3.60±0.10[d]
V160A	1.232±0.008	2.10±0.05	+2.46±0.09	+1.91±0.04

Global stability differences (U↔N) between the protein variants and WT.

[a] Errors calculated by the method of Bevington, as described.[50] The m values are the constants that relate the concentrations of mid denaturation of the different proteins with the corresponding free energy differences between the native and the denatured states (equation (5)).

[b] Global (U↔N) free energies of folding of mutants minus that of WT, calculated as $m(\text{mut})U_{1/2}(\text{mut}) - m(\text{WT})U_{1/2}(\text{WT})$. Errors propagated from the corresponding m and $U_{1/2}$ values.

[c] Global (U↔N) free energies of folding of mutants minus that of WT, calculated as $m_{av} \times \Delta U_{1/2}$, where m_{av} is the mean value of the m slope of the different proteins (2.282±0.045, mean±SE). Errors propagated from that of m_{av} and those of the corresponding $U_{1/2}$ values. The $\Delta\Delta G_{fold}$ calculated by this procedure are more accurate than those calculated as explained in the previous footnote.

[d] The m values of these two mutants are not within the SD of the other m values and it is not appropriate to use an average m value to calculate their folding free energy differences. Therefore, the free energy differences reported are those calculated as explained in footnote **b**.

from the calculated stability changes (see below). To overcome this limitation we use now an alternative approach. Both $\Delta\Delta G_1$ and $\Delta\Delta G_2$ are derived from wild-type and mutant ΔG values calculated using equation (1) at a single reference temperature that is carefully chosen so that the neglected contributions of the ΔC_p terms in equations (7) and (8) are at a minimum. To select the reference temperature, $\Delta\Delta G_1$ and $\Delta\Delta G_2$ values are calculated for each mutant as a function of temperature using either equations (7) and (8) or the simplified equation (1). From either of the two sets of $\Delta\Delta G_1$ and $\Delta\Delta G_2$ values, φ-values are calculated (see equation (3) below) as a function of temperature. The difference between the two sets of φ-values is an estimation of the errors introduced by having removed, in equation (1), the ΔC_p term. Each mutant shows a temperature where the two φ-values are identical. Since a single reference temperature is required, the average of those temperatures has been initially selected as the common reference temperature to perform the φ-analysis. Unlike what could have been expected, the reference temperature is not close to the average of the T_{m1} and T_{m2} values of the different proteins. The fact that most mutants are destabilising and that ΔC_{p2} is smaller than ΔC_{p1} sets the reference temperature (316.2 K) very close to the first melting temperature of the wild-type protein (317.3 K). For simplicity, we have selected 317.3 K as the reference temperature to calculate the changes in free energy of the two equilibria that are shown in Table 3.

The accuracy of the extrapolation of the thermal unfolding data to the reference temperature has been tested experimentally by recording the urea-denaturation curve of the wild-type protein at

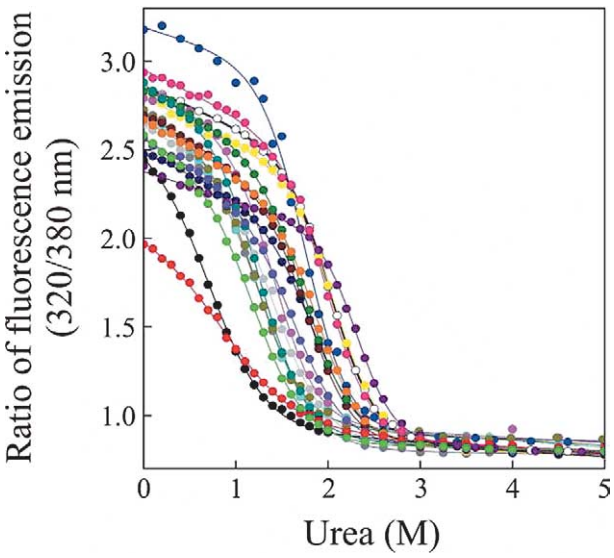

Figure 3. Urea-denaturation curves of wild-type (open circles), L6A (black), I22V (light grey), V31A (dark grey), D43A (yellow), I51V (cyan), I52V (magenta), S71A (blue), A84G (dark yellow), N97A (dark blue), Q99A (purple), I104V (wine), S110A (olive), V117A (dark cyan), T122S (orange), V139A (violet), L143A (pink), I156V (red) and V160A (green). The lines are fits to a two-state equation. The curves were acquired at 25.0 °C in 50 mM Mops (pH 7.0).

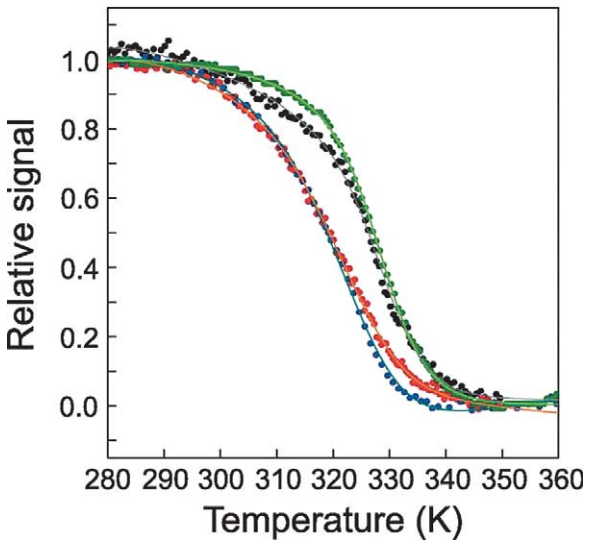

Figure 4. Global fitting (continuous lines) of experimental thermal denaturation curves of the wild-type protein followed by far-UV CD (black), near-UV CD (red), fluorescence (blue) and absorbance (green). Each curve was normalized to values from roughly 1 to 0 before the fitting. The buffer used was 50 mM Mops (pH 7.0).

317.3 K (Figure 5(a)). At this temperature, the native and intermediate states of the wild-type protein are equally populated (while the concentration of denatured state is negligible) and adding urea promotes the unfolding of both conformations. The stability of the intermediate at this temperature (ΔG_{ID}), calculated from a fit of the urea-unfolding curve (Figure 5(a)) to equation (9), is of 1.61($\pm$0.17) kcal mol^{-1}, which agrees well with the value calculated from the global fit of the thermal unfolding curves (see fitted data in Table 3) when it is extrapolated to the reference temperature of 317.3 K using equation (1) (1.85($\pm$0.16) kcal mol^{-1}). In the fit, the *m* value of the native state (proportionality constant between free energy difference and urea concentration) is assumed to be that measured previously at 25.0 °C. The *m* fitted value for the intermediate is a bit lower, as expected (not shown).

An additional test of the feasibility of using equation (1) to extrapolate the thermal unfolding data to the common reference temperature selected comes from a comparison of the global stability differences ($\Delta\Delta G_{ND}$, wild-type − mutant) determined by adding the $\Delta\Delta G_{NI}$ and $\Delta\Delta G_{ID}$ values extrapolated to that temperature (data in Table 3) with those directly determined at 298.2 K using urea-unfolding curves (data in Table 2). The two data sets need coincide only if the thermally and urea-denatured states are energetically equivalent, if the stability differences between proteins are retained on going from 317.3 K to 298.2 K, and if the use of the simplified equation (1) is appropriate. In fact, the correlation between the $\Delta\Delta G_{ND}$ values calculated by either procedure (Figure 5(b)) is quite good: $\Delta\Delta G_{ND}^{thermal} = -0.2 + 1.06 \times \Delta\Delta G_{ND}^{urea}$; $R = 0.91$ ($\Delta\Delta G_{ND}^{urea}$ values calculated using an average *m* value, see Table 2; the two mutants with divergent *m* values (see legend) are excluded from the fit) and thus unlikely to arise from a compensation of errors see Campos *et al.*,[31] for a similarly good correlation applying to a different set of mutants).

ϕ-Analysis of the equilibrium intermediate

By analogy with the ϕ-analysis of transition states of protein folding and of transient folding intermediates,[25] a parameter can be defined to describe the integrity in the thermal intermediate of native side-chain interactions probed by mutation (equation (3)):

$$\phi_{int} = \frac{\Delta\Delta G_2}{\Delta\Delta G_1 + \Delta\Delta G_2} \quad (3)$$

where the subscripts 1 and 2 refer to the N-to-I and the I-to-D transitions, respectively. The integrity parameter takes a value of $\phi = 0$ when the free energy difference of the second transition is not modified by the mutation (i.e. when the interaction probed is not present in the intermediate), and takes a value of 1 when the mutation only modifies the free energy differences of the second transition (i.e. when the interaction probed is similarly strong in the native state and in the intermediate). The interpretation of fractional values of the integrity parameter is subjected to the same considerations as in classic ϕ-analysis.[25]

In a previous work,[28] we defined an equilibrium ϕ-value in a similar way, but the $\Delta\Delta G_1$ and $\Delta\Delta G_2$ energies used in its calculation were each determined at the transition temperature of the corresponding equilibria and, therefore, the ϕ-values could not be ascribed to a defined temperature. Here, we have chosen a common temperature of reference to calculate the two energy differences. As explained above, the temperature has been carefully selected so that long extrapolations from the transition temperatures are avoided and the error introduced by using the simplified equation (1) is minimized. The ϕ-values calculated in this way (Table 3) describe the structure of the thermal intermediate at a defined temperature: 317.3 K. They are all between 0 and 1, except those corresponding to D43A, S71A, which are unreliable due to the small energy changes involved.

Discussion

Equilibrium ϕ-analysis of stable protein folding intermediates

The main goal in this work is to obtain a low-resolution structure of the apoflavodoxin thermal intermediate. To that end we have developed a general method of equilibrium ϕ-analysis that can be applied to investigate the structure of

Table 3. Thermal unfolding of wild-type and mutant apoflavodoxins

Protein	N-to-I equilibrium[a]		I-to-D equilibrium[a]		Differential energies[b]		
	ΔH_1 (kcal mol^{-1})	T_{m1} (K)	ΔH_2 (kcal mol^{-1})	T_{m2} (K)	$\Delta\Delta G_1$ (kcal mol^{-1})	$\Delta\Delta G_2$ (kcal mol^{-1})	ϕ[c]
Wt	33.9±0.1	317.3±0.2	52.7±0.6	329.0±0.1	–	–	–
L6A	28.7±0.8	305.5±0.2	23.8±0.7	319.1±0.2	1.11±0.04	1.74±0.03	0.61±0.01
I22V	33.0±0.8	311.0±0.2	37.3±1.2	324.7±0.3	0.67±0.03	1.03±0.04	0.61±0.04
V31A	30.9±0.1	312.6±0.2	41.5±0.8	324.2±0.2	0.46±0.02	0.99±0.04	0.68±0.03
D43A	34.1±0.2	317.7±0.4	49.5±1.5	329.6±0.2	−0.04±0.04	0.03±0.06	−1.68±9.5
I51V	31.2±0.1	313.2±0.2	38.8±0.9	324.3±0.2	0.41±0.02	1.04±0.03	0.72±0.04
I52V	39.8±0.3	315.9±0.1	54.5±1.2	324.9±0.1	0.18±0.02	0.60±0.04	0.77±0.07
S71A	34.5±0.1	316.3±0.2	55.0±0.8	329.3±0.1	0.11±0.02	−0.13±0.04	6.18±14.2
A84G	34.4±0.2	313.0±0.1	31.9±0.7	323.6±0.1	0.47±0.02	1.25±0.03	0.73±0.02
N97A	28.6±0.2	314.9±0.4	48.7±0.6	328.3±0.1	0.22±0.04	0.24±0.03	0.53±0.10
Q99A	52.7±0.6	322.5±0.1	56.2±1.2	331.3±0.1	−0.85±0.02	−0.49±0.06	0.37±0.05
I104V	29.5±0.2	313.3±0.3	54.5±0.8	327.9±0.1	0.38±0.04	0.11±0.04	0.23±0.09
S110A	32.5±0.2	315.6±0.4	44.3±1.5	328.7±0.2	0.18±0.04	0.34±0.06	0.66±0.16
V117A	36.9±0.7	309.3±0.1	33.7±0.7	322.6±0.1	0.95±0.03	1.32±0.03	0.58±0.02
T122S	35.5±0.2	323.2±0.4	50.7±1.5	330.3±0.1	−0.65±0.04	−0.11±0.07	0.16±0.09
V139A	29.0±0.1	310.9±0.2	50.8±0.6	327.5±0.1	0.60±0.02	0.29±0.03	0.33±0.04
L143A	33.3±0.1	315.8±0.2	50.0±0.5	328.9±0.1	0.15±0.02	0.11±0.03	0.41±0.14
I156V	25.3±0.2	308.8±0.4	29.9±0.9	320.3±0.2	0.69±0.04	1.60±0.03	0.70±0.02
V160A	27.7±0.3	309.1±0.6	49.2±1.4	322.4±0.2	0.73±0.05	1.10±0.04	0.60±0.03

Partial stability differences ($\Delta\Delta G_1$ and $\Delta\Delta G_2$) between mutant proteins and wild-type, and ϕ-values of the probed interactions at 317.3 K.

[a] The $\Delta H(T_m)$ and T_m values of the two equilibria are derived, for each protein, from a global fit of four thermal unfolding curves recorded using fluorescence emission, near-UV CD, near-UV absorbance, and far-UV CD. The standard deviations have been calculated by interval analysis.[50] Reducing to 1/20th of the data points the data sets of each protein that are globally fitted to the three-state equations leads to SD for the T_m and ΔH values about two and three times larger, respectively, than those reported in the Table. The propagation of those SD to the $\Delta\Delta G$ values approximately doubles their SD.

[b] Unfolding free energy differences (wild-type−mutant) calculated at 317.3 K (see Methods). Errors propagated from those of $\Delta H(T_m)$ and T_m.

[c] Integrity of the probed interaction in the thermal intermediate. Calculated, at 317.3 K, as $\Delta\Delta G_2/(\Delta\Delta G_1+\Delta\Delta G_2)$. Errors propagated from those of $\Delta\Delta G_1$ and $\Delta\Delta G_2$.

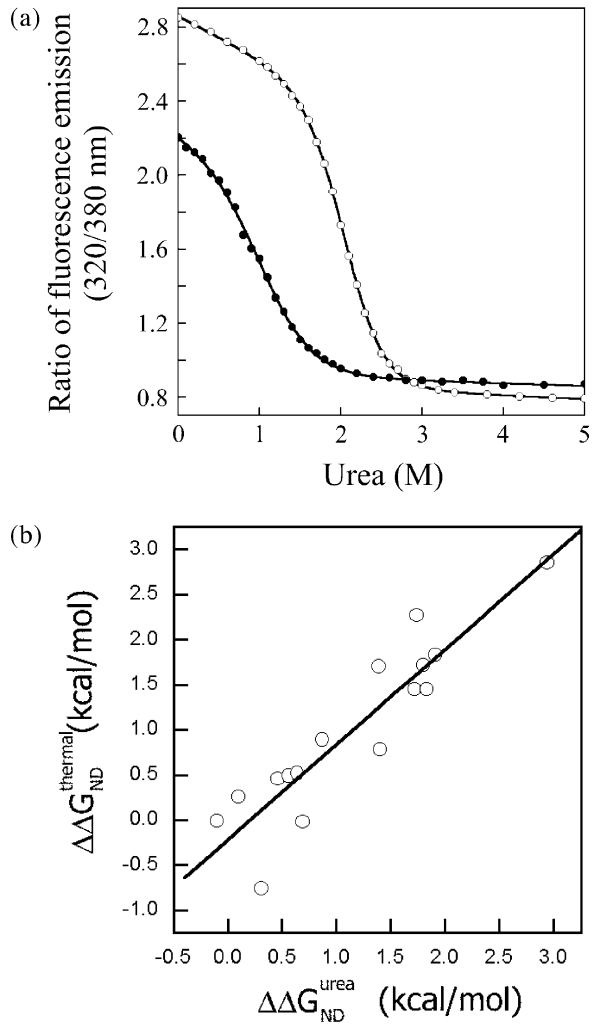

Figure 5. Comparison of thermal and urea-denaturation analyses. (a) Comparison of urea-denaturation curves of wild-type apoflavodoxin measured at 25 °C (open circles) and at 44 °C (filled circles). The fitted curves are shown as continuous lines. The fitting of the 44 °C curve (where the native and intermediate states are present at equimolar concentrations) allows us to calculate the free energy difference of the I-to-D equilibrium (see Methods) and to compare this value with that obtained from the global fitting of the thermal unfolding data. (b) Comparison of $\Delta\Delta G_{ND}$ values obtained by adding the $\Delta\Delta G_{NI}$ and $\Delta\Delta G_{ID}$ terms calculated from global analysis of three-state thermal unfolding curves (data in Table 3) with those directly measured from two-state urea-denaturation.

the thermal intermediates of other proteins. Initially, φ-analysis was conceived as a tool to investigate transition states of protein unfolding. The essence of the method, however, is its ability to derive structural information on a protein conformation that is related to two additional ones (one of known structure and the other one assumed to be disordered) by measurable equilibria that can be perturbed by mutation. This allows the method to be extended to investigate equilibrium intermediates. In fact, although we have used the method to investigate an intermediate of thermal unfolding, it could also be adapted to study the structure of other types of intermediates (such as intermediates at mildly acidic pH values) provided that the free energy differences of the two equilibria can be measured accurately.

The φ-values derived for the 18 mutants specifically produced to obtain the structure of the apoflavodoxin intermediate are shown in Table 4. Additionally, we have made use of the φ-values calculated for other mutants that are being studied in the laboratory for various reasons and whose full thermodynamic analysis will be published elsewhere. These additional mutants are I21G, G68A, E107G and Q111A (located in α-helices), L50A and L105A (that make cavities in the protein), and E20K, E40K, E72K, D75K, D126K and D150K,[31] and their φ-values have been calculated in the same manner. Furthermore, we have re-analysed one surface-exposed hydrogen bond-breaking mutant already published[28] (Y47F) with the improved method used here that calculates the φ-values at a well-defined temperature. Of the 31 mutants analysed, 29 show ΔG_{ND} values that are large enough so that the errors in the calculated ϕ_{int} values are small. Two of them, nevertheless, D43A and S71A, do not fulfil this premise, and have not been used for the structural analysis.

The errors reported for the φ-values ultimately depend on the errors associated with the transition temperatures from which they have been calculated by propagation. We notice that we have not used for these primary errors the values provided by the fitting program. Instead, they have been obtained from a global confidence analysis of the different thermal unfolding curves available for each mutant, as described in Methods. Since the errors obtained in this type of analysis depend on the number of data points used, we have additionally examined what effect would have reducing the number of experimental data points in the global fit performed for each mutant to just 1/20th of the available points. As expected, the errors calculated in this way are larger (not shown), but they only double the errors reported for the φ-values in Table 3, which are still low. Besides, we have determined experimental errors for the transition temperatures of the wild-type protein by performing four independent fits of independent sets of thermal unfolding curves. The standard error calculated for T_{m2} is 0.19, similar to those derived from the confidence analysis performed for the different proteins (Table 3). The standard error of T_{m1} is a bit larger: 0.61. Assuming these errors are representative for the variants, the propagation gives for the φ-values errors that are 1.5–2 times larger than those reported in Table 3, still low compared to the φ-values.

On the other hand, there has been concern in the literature on when $\Delta\Delta G_{ND}$ values are large enough so that φ-values are valid. Recently,[35] it has been suggested that they should be greater than

1.7 kcal mol^{-1}. However, when the data set used to analyse the issue was restricted to non-disruptive mutations, ϕ-values above 0.6 kcal mol^{-1} seemed to be appropriate.[36] The mutants presented in this work belong to this second class (Table 1). In our analysis we use the stability data derived from thermal unfolding analysis to calculate equilibrium ϕ-values (the urea-denaturation data being just confirmative of the quality of the thermal analysis). It is thus interesting to see how the $\Delta\Delta G_{ND}$ data in Table 3, compare with the above 0.6 kcal mol^{-1} cut-off criterion. As already discussed, two of the 31 mutants analysed (D43A and S71A) give rise to negligible free energy changes ($\Delta\Delta G_{ND}$ of -0.01 and -0.02) and thus to very large errors in the corresponding ϕ-values. These two mutants have not been used to delineate the ϕ-structure of the intermediate. Of the 29 remaining mutants, 22 display free energy changes that are larger than the 0.6 kcal mol^{-1} cut-off suggested,[36] while seven mutants show smaller but significant changes (from 0.26 kcal mol^{-1} to 0.52 kcal mol^{-1}). Even if we do not consider these seven mutants, the structure of the apoflavodoxin thermal intermediate would look the same: in other words, the ϕ-values calculated to the seven mutants exhibiting an average $\Delta\Delta G_{ND}$ value of 0.43 kcal mol^{-1} are fully consistent with the structure based in the ϕ-values corresponding to the 22 mutants with free energy differences larger than 0.6 kcal mol^{-1}. Since, in addition, the structure calculated (Figure 6(a)) from the ϕ-analysis is fully consistent with the NMR data available (see below and Figure 6(b)), it seems that equilibrium ϕ-values calculated from thermal unfolding data carefully analysed, as described in Methods, and corresponding to non-disruptive mutations, can be trusted if the corresponding $\Delta\Delta G_{ND}$ values are greater than 0.4 kcal mol^{-1}.

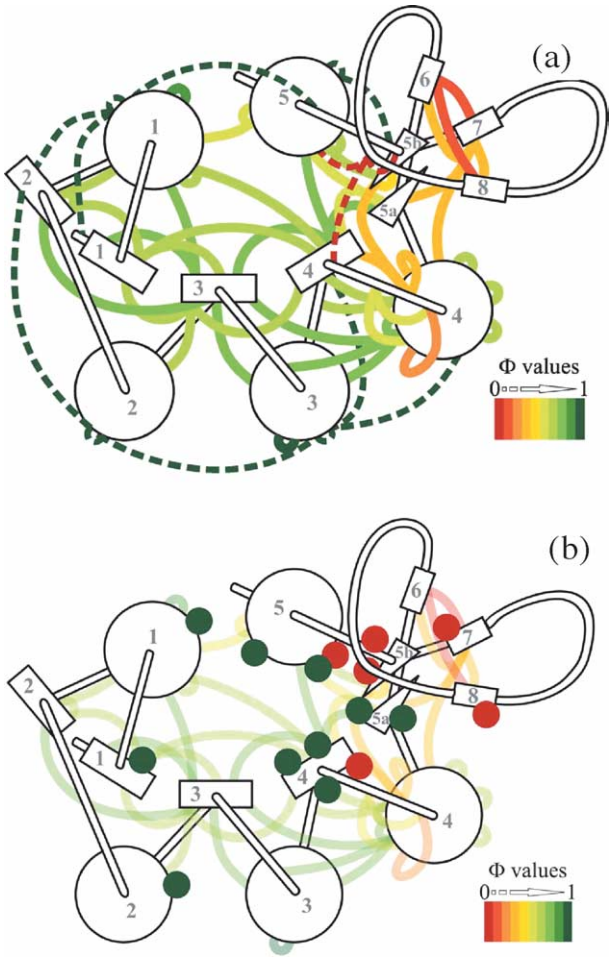

Figure 6. Scheme of the structure of the apoflavodoxin thermal intermediate. (a) ϕ-Structure. The ϕ-values are shown in a colour code: totally formed (dark green); totally broken (red). Furthermore, ϕ-values of several Asp or Glu to Lys mutants (E20K, E40K, E72K, D75K, D126K and D150K[31]) are shown as broken lines using the same color code. (b) NMR data gathered on the E20K/E72K apoflavodoxin mutant whose thermal intermediate constitutes 83% of the protein molecules at 55.4 °C (not shown). Green dots represent the position in the structure of protons displaying native chemical shift values at 55.4 °C. Red dots represent protons whose chemical shifts in the native state at 20 °C markedly deviate from typical random coil values but that, at 55.4 °C, present random coil chemical shifts.

The structure of the apoflavodoxin thermal intermediate at 317.3 K

Among the mutations analysed, ϕ-values of 0 and 1 are scarce, except for some electrostatic interactions. This probably reflects that, even in the more structured regions of the intermediate, the native interactions are somewhat weakened and that in the less ordered regions some very debilitated interactions may persist. The overall picture of the structure of the intermediate is nevertheless very clear and it is shown in Figure 6, using a colour code to facilitate the visual interpretation of the ϕ-values, which are represented as lines connecting the interactions broken upon mutation. When a given mutation disrupts more than one interaction, as it is typical of mutations in the core of the protein, one line has been drawn for each broken interaction. By doing so, we favour an interpretation of those fractional ϕ-values as arising from the simultaneous debilitation of the two interactions involved rather that from the full conservation of the interaction with one neighbour and the full destruction of the additional interaction. We think that this makes sense in a protein core, specially when so clear a bipartite structure of the intermediate becomes evident from a simple visual inspection of Figure 6. As it has been discussed,[25] fractional ϕ-values can also arise from conformational heterogeneity. Although this possibility can rarely be fully discarded, we notice that the thermal unfolding of the intermediate is quite cooperative, as it occurs in a short temperature range and with a large enthalpy change, which argues against any marked conformational heterogeneity.

As shown in Figure 6, the structure of the apoflavodoxin thermal intermediate at 317.3 K is

mainly formed by the packing of helices 1, 2, 3, 4 and 5, and strands 1, 2, 3 and 4 (where the φ-values are higher than 0.6) and possibly by strands 5a and 5b (that seem weakened, see also NMR evidence). In contrast, the long loop splitting strands 5a and 5b (that contains a small three-stranded β-sheet comprising strands 6, 7 and 8), and the loop connecting strand 4 and helix 4 are markedly weakened (all φ-values below 0.4). The intermediate thus contains the two original hydrophobic cores formed by the packing of the α-helices onto the central β-sheet, while the 90 to 100 and 120 to 139 loops are disordered. We notice that this structural partition of the native structure into two different regions in the intermediate has been inferred without making consideration of the φ-values derived from several charge reversal stabilising mutations at the surface of the protein (E20K, E40K, E72K, D75K, D126K and D150K)[31] that are represented in Figure 6 as broken lines, with the same colour code as the other φ-values. We have chosen to distinguish these mutations from the others because their effects on the stability are exerted, in principle, through many charge/charge interactions all over the protein. Electrostatic considerations, nevertheless, allow us to identify the few residues that are strongly coupled with each of the replaced ones. Five of those six charge reversal mutations are located in helices 1, 2 and 3, and extend their influences into helices 4 and 5, and into strands 1, 2 and 4. They are thus placed in and interacting mainly within the region of the intermediate that has been defined as close to natively structured. Not surprisingly, their φ-values are high, in fact very close to 1. In contrast, the charge reversal mutation D126K is located in the region delineated as being poorly structured, to which the stronger electrostatic interactions of the charged residue are confined. Significantly, its φ-value is close to 0, which indicates that, in the intermediate, it becomes uncoupled from the structured region. We have additionally studied a charge reversal mutation, E61K, that is located in one of the FMN-binding loops of the protein: the loop connecting strand 3 with helix 3. Although its φ-value is strange (negative, not shown) due to the fact that it exerts opposite effects on the free energy of the two equilibria, it clearly stabilises the native state relative to the intermediate, which is consistent with its presence in a region where the native and intermediate states differ significantly. Based on this fact, and taking also into account that the 57–63 loop has already been shown to present a large plasticity even at 288.2 K in both solution[30] and different crystals,[37] we propose that the loop is disordered in the thermal intermediate. Our mutational analysis thus indicates that the regions unstructured in the thermal intermediate correspond essentially (see Figure 7(a)) to three segments of sequence (57 to 63, 90 to 100, and 120 to 139 that are in contact in the 3D structure (see Figure 7(a)). It is clear that residues located in the boundary between the native-like and the non-structured regions of the intermediate may be somewhat rearranged.

NMR confirmation of the φ-structure of the apoflavodoxin intermediate

To test this method of equilibrium φ-analysis, we have obtained some fully independent, preliminary information on the structure of the thermal intermediate by using NMR. Complications here arise from the fact that, according to our global analysis, at the temperature of highest accumulation, the intermediate only constitutes a mere 64% of the protein population, the rest being a mixture of fully native and fully unfolded protein. To overcome this limitation, we have combined two stabilising mutations (E20K and E72K[31]) to produce an apoflavodoxin mutant whose thermal intermediate accumulates up to 83% at 55.4 °C (L.A.C. & J.S., unpublished results). Using this mutant, we have recorded 1D proton NMR spectra where distinct high-field signals assigned to Leu141 and Leu163 protons have been followed over the N-to-I thermal transition (Figure 8). Two peaks report on an interaction between the Leu141 side-chain and an aromatic residue (W120) located in the unstructured region of the intermediate, while an additional peak reports on an interaction between Leu163 and Phe26, both residues located in the native-like region of the intermediate. The 1D proton spectra reveal that, while the Leu141 signals fade away at 55.4 °C, that of Leu163 does not, in full agreement with the structure proposed for the thermal intermediate from φ-analysis.

In addition, we have performed, on the basis of the assignment reported for the wild-type apoflavodoxin,[30] a partial assignment of proton resonances in the native state of the E20K/E72K mutant at 20 °C. By following signals in 2D-TOCSY and 2D-NOESY spectra recorded at different temperatures from 20 °C to 55.4 °C we have identified two sets of signals with very different behaviour. One set was almost unaffected by temperature increase from 20 °C to 55.4 °C (see Supplementary Material), which indicates that the corresponding residues display the same structure in the native state and in the intermediate. The other set of signals exhibits δ-values that, in the native state, deviate strongly from random coil values (this facilitates their assignment in the native state of the mutant by comparison with the assignment of wild-type protein) but that shift towards random coil values in the intermediate, which clearly indicates that they no longer retain their native environment in the intermediate conformation. The correspondence between the residues displaying native state δ-values at 55.4 °C and the regions determined to be close to natively folded in the φ-structure of the intermediate is excellent (Figure 6(b)), as it is for the residues with random coil values at 55.4 °C and the unstructured regions of the intermediate.

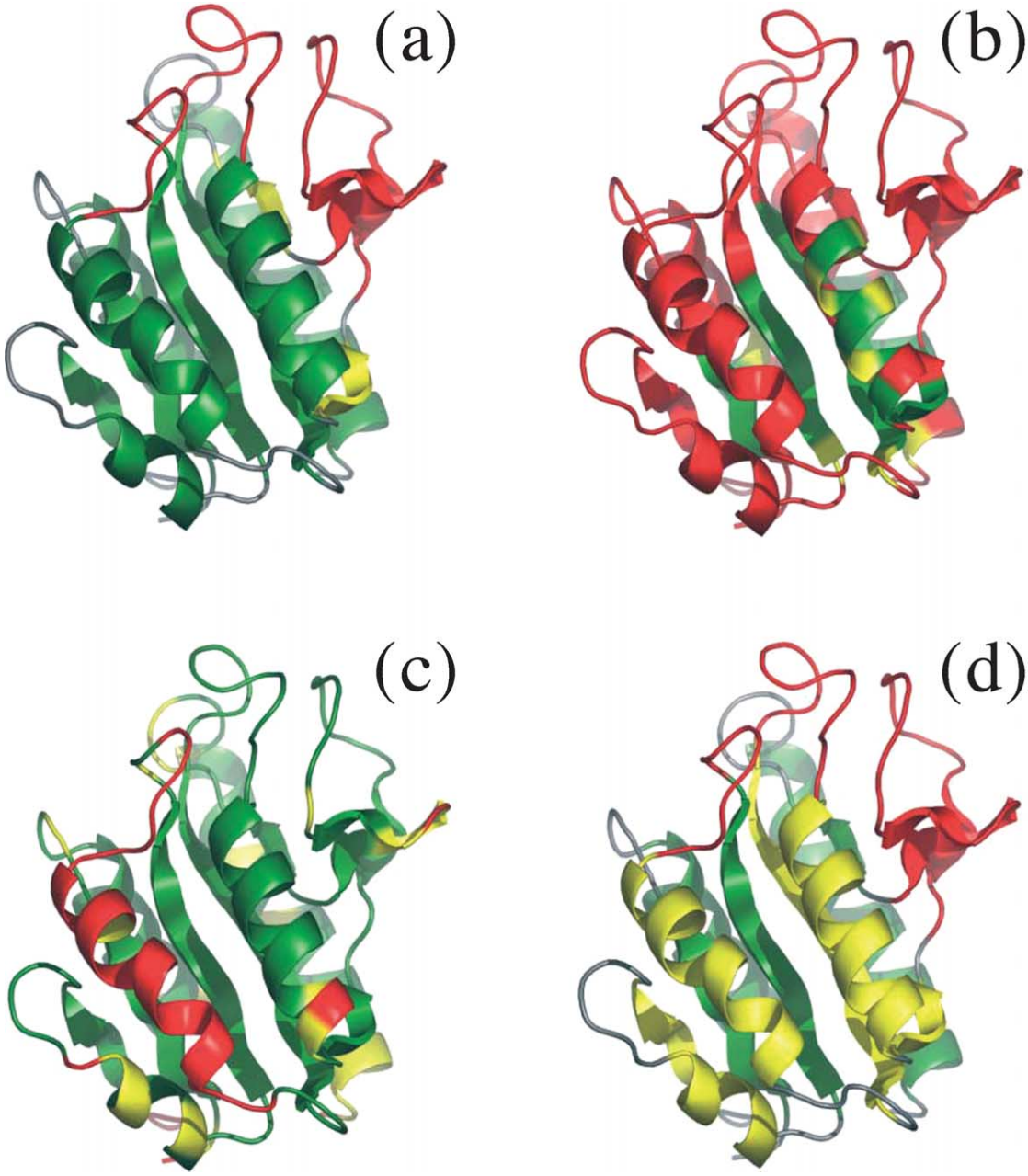

Figure 7. Comparison of (a) the structure of the apoflavodoxin thermal intermediate (this work) and (b) the proton slow-exchange core of the protein previously reported.[30] In addition, the native structure is represented coloured according to (c) the *B*-factor of the C_α[29] and (d) the percentage of apolar surface buried between a given structural element in its native conformation and the rest of the protein. The colour code used is red (for regions unstructured in the intermediate, as defined by ϕ-analysis; or local ΔG opening[30] <5.5 kcal mol^{-1}; or *B*-factor >20; or apolar ΔSAS $<64\%$); yellow (regions that could be unstructured in the intermediate; or local ΔG opening between 5.5 kcal mol^{-1} and 6.5 kcal mol^{-1}; or *B*-factor between 10 and 20; or apolar ΔSAS between 64 and 71%); and green (close to native regions of the intermediate; or local ΔG opening >6.5 kcal mol^{-1}; or *B*-factor <10; or apolar ΔSAS $>71\%$). The structured regions of the apoflavodoxin thermal intermediate do not correspond with those of the experimentally determined slow-exchange core or those displaying high *B*-factors. They roughly correspond with the structural elements whose packing with the rest of the protein buries a higher percentage of apolar area.

This preliminary NMR analysis of the structure of the apoflavodoxin intermediate, using the E20K/E72K double mutant as a model, strongly supports the equilibrium ϕ-analysis structure and additionally clarifies the status of the split fifth β-strand: β5a seems folded, while β5b appears unfolded in the intermediate.

Can the structure of the thermal intermediate be anticipated from hydrogen exchange analysis or by some other criteria?

The structure of the intermediate (Figure 7(a)) reveals a distinct region of apoflavodoxin that is markedly unfolded and a larger one that is more

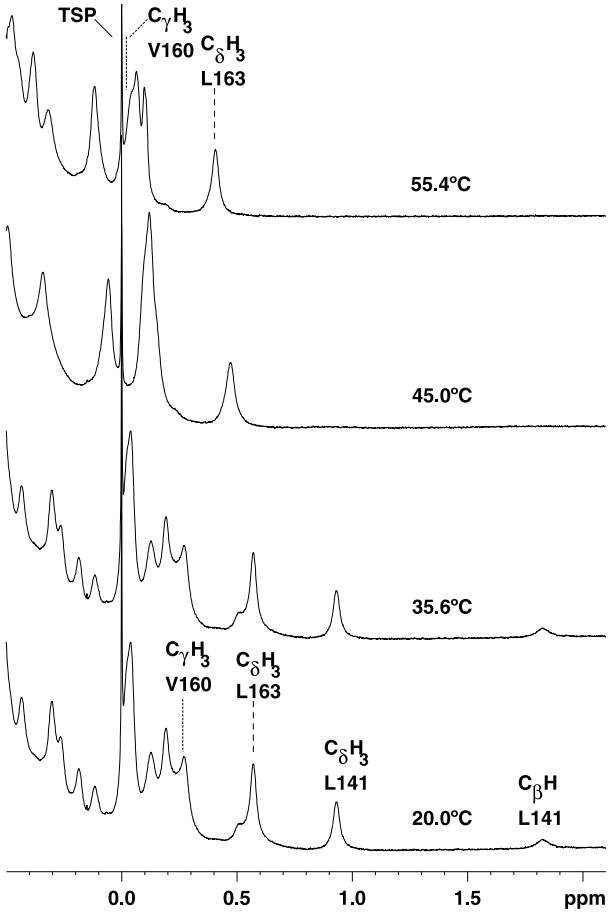

Figure 8. One-dimensional proton NMR spectra (high-field region) of the E20K/E72K-apoflavodoxin mutant at different temperatures, as indicated. The protein sample (0.5 mM) was dissolved in 25 mM sodium phosphate at pH 7.0. Under these solution conditions, the maximal accumulation of intermediate (83%) is greater than that of the wild-type protein and takes place at 55.4 °C (unpublished results). Distinct high-field signals, corresponding to Leu141 and Leu163 protons can be followed over the N-to-I thermal transition. Two peaks report on an interaction between the Leu141 side-chain with an aromatic residue (Trp120) in the unstructured region of the intermediate, while an additional peak reports on an interaction between Leu163 and Phe26, both in the native-like region of the intermediate. Besides, one peak corresponding to Val160 (located in the native-like region) is also shown, whose chemical shift at 55.4 °C has been assigned from 2D spectra (see Table 4). The 1D proton spectra reveal that, while the Leu141 signals fade away at 55.4 °C, those of Leu163 and V160 do not, in full agreement with the structure proposed for the thermal intermediate from φ-analysis.

structured. Similar behaviour has been described for other protein equilibrium intermediates studied by NMR.[38,39] It would be most useful if the structure of a protein intermediate could be anticipated from an analysis of native state properties. Experimentally, the regions of a native protein that are less likely to experience partial transient unfolding events can be determined from NMR studies that monitor the speed at which the acidic protons of the protein are exchanged with solvent protons.[40–42] In this way, the slow-exchange stability cores of a number of proteins have been determined.[43] Educated guessing would tend to identify the regions that more actively exchange protons with solvent as those more likely to become denatured when the native state is destabilised (say, by raising the temperature) and an intermediate conformation, not fully denatured, appears. This guess can be easily tested in apoflavodoxin because its native-state hydrogen exchange has been previously determined (at 25 °C).[30] A simple inspection of Figure 7(a) and (b), where the structures of the thermal intermediate (this work) and that of the slow-exchange core[30] are compared, shows that they do not coincide. One possible explanation for this lack of correspondence may be found in the fact that the hydrogen-exchange experiments were performed at 298.1 K, while the structure of the intermediate has been determined at 317.3 K (a further extrapolation of the stability data presented in this work to 298.1 K would be unwise and it is not attempted). At the higher temperature where the intermediate accumulates, the hydrophobic effect is strengthened and the hydrophobic cores of the protein may become more resistant to unfolding thus making the structure of the thermal intermediate divergent from that of the slow exchange core.

Alternatively, the differences in the two structures may simply reflect the fact that the exchange rates of a large part of the apoflavodoxin amide protons (as in other proteins) are too fast to be measured accurately.[30] Since this applies to both regions that remain structured and regions that are denatured in the intermediate, the possibility cannot be ruled out that the regions unfolded do show higher exchange rates than the rest. Even if that were the case, it is clear that the structure of the thermal equilibrium intermediate does not correspond to that of the slow-exchange core as it can be experimentally determined (although the intermediate includes the slow-exchange core). Therefore, knowing the slow-exchange core of apoflavodoxin does not help to predict the structure of its intermediate. Whether this applies to other proteins remains to be tested.

We have thus investigated other possible ways to predict, from native state properties, the structure of the intermediate. One possibility is that the unstructured regions of the intermediate correlate with the segments of the protein with higher *B*-factors in the X-ray structure.[29] However, the *B*-factors of the C_α clearly show (Figure 7(c)) that the greatest mobility appears in the third helix and in the loop between strand 3 and helix 3, while the long loop 120–139 displays low *B*-factors. We must thus conclude that the structure of the intermediate is unrelated to the *B*-factors of the crystal structure. Finally, we have explored a possible relationship between the regions that are disordered in the thermal intermediate and those whose interface with the rest of the protein in the native structure is more polar. To that end, we have calculated the

solvent-accessible surfaces (SAS) of each secondary structure element and main loop in apoflavodoxin (and of the corresponding complementary peptides) in hypothetical isolation, but retaining the native conformation. Then, for each pair of complementary fragments, the percentage of buried area that corresponds to apolar atoms has been calculated. This is represented, colour coded, in Figure 7(d). The three loops unfolded in the intermediate (one including a small β-sheet) are indeed, when compared to any of the helices or β-strands of the central β-sheet, the structural elements where the ratio of apolar/total area buried is lower. We see thus qualitative agreement between high polar surface burial of individual structural elements in the native state and their lack of structure in the thermal intermediate. We believe this correlation is worth testing in other proteins to see whether the unstructured regions of protein thermal intermediates can be predicted.

The structure of the thermal intermediate can guide the selective stabilisation of distinct conformations of a protein

A final issue in this work is related to the rational modification of energy barriers between folded, intermediate and unfolded states. Over the years, many mutational strategies have been proposed that, in many cases, have succeeded in increasing the energy difference between the folded and unfolded states of a given protein. In most cases, fairly small proteins with simple two-state equilibrium behaviour have been used to demonstrate the feasibility of a particular approach. It seems important to realise, however, that most proteins are likely to display more complex equilibrium behaviours. Examples of equilibrium three-state proteins abound.[4,8,28,38,44] In one recent work, the unfolding equilibrium of a single chain Fv antibody fragment was described as four-state.[5] For any protein displaying non-functional equilibrium intermediates between the native and unfolded states the energy difference that confines the polypeptide conformation into the active conformation is the free energy difference between the native state and the first non-functional intermediate. This free energy difference has been termed the relevant stability of the protein,[45] as opposed to the residual stability of the intermediate (relative to the unfolded state). From the work done by Campos *et al.*,[31] we determine, for a set of charge reversal mutations at the surface of apoflavodoxin, the overall, relevant and residual stability of the protein. We find that, although most mutations tested markedly increase the overall stability of the protein, only a fraction of them realise this increase into relevant stabilisation. Most stabilising mutations simply widen the energy gap between the thermal intermediate and the unfolded state. When the two kinds of mutations are mapped onto the structure of the thermal intermediate they clearly segregate, so that while the mutations that increase the relevant stability appear in the unstructured regions of the intermediate, those increasing the residual stability are all located in the structured region. Although perhaps not surprising, this clustering nicely demonstrates a couple of things. First, that the strategies developed over the years to increase the energy gaps between folded and unfolded conformations may not be similarly successful in stabilising native conformations against equilibrium intermediates that resemble the native conformation. Second, that knowledge of the structure of an intermediate at the residue level can efficiently guide the rational increase of the relevant stability of complex proteins by allowing us to concentrate efforts in the more promising regions. In this respect, the method of φ-analysis of thermal intermediates developed here to investigate the apoflavodoxin thermal intermediate can be of much help. Obviously, the target regions for increasing the relevant stability of a protein by site-directed mutagenesis essentially coincide with the regions where that stability can be increased through ligand binding. Knowing the structure of an equilibrium intermediate can thus be also useful to attempt the selective stabilisation of native states by the latter approach. Finally, we notice that, as illustrated in this work, the tailoring of the relevant and residual stabilities of a protein (once the φ-structure of its intermediate has been determined) allows us to design protein variants with intermediates that, at a selected temperature, constitute the largely dominant conformation in solution, which can in turn facilitate a further structural characterization of the intermediate by NMR.

Methods

Design of mutants

Mutations (to shorter residues, with branching conservation, and excluding large deletions) reporting on interactions between secondary structure elements have been preferred over those reporting on local interactions (see Table 1). In addition to those of the 18 specifically designed mutants, the φ-values of several positions mutated in studies in progress in the laboratory (M.B., J.L.-L. & J.S., unpublished results) and of one position already reported[28] have been used to delineate the structure of the apoflavodoxin thermal intermediate. Thirty-one mutants have been analysed.

Mutagenesis, protein expression and purification

PCR mutagenesis of the selected mutants was performed using the protocol described in the Stratagene QuikChange kit. The mutations were identified by sequencing the entire gene[46] and the mutants were expressed, purified and the FMN cofactor removed as described.[33]

Circular dichroism spectra

Far-UV spectra were measured from 200 nm to 250 nm in a 1 mm path-length cuvette, with samples of 15 μM protein in 5 mM Mops (pH 7) with 15 mM NaCl. This

$$Y = \frac{Y_N + m_N T + (F_I + m_I T)e^{-\Delta G_1/RT} + (F_D + m_D T)e^{-(\Delta G_1 + \Delta G_2)/RT}}{1 + e^{-\Delta G_1/RT} + e^{-(\Delta G_1 + \Delta G_2)/RT}} \quad (6)$$

buffer is of the same ionic strength, but less absorbance, as the 50 mM Mops buffer used for most of the experiments. Near-UV spectra were recorded from 260 nm to 310 nm in a 1 cm path-length cuvette with samples of 30 μM protein in 50 mM Mops (pH 7). All CD spectra were recorded in a JASCO 710 spectropolarimeter, thermostated at 25.0(±0.1)°C, using a scan rate of 20 nm/minute.

Urea denaturation curves

Protein solutions (2 μM) of different urea concentrations, from 0 M to 5 M, were prepared and their ratios of fluorescence emission at 320 nm and 380 nm, with excitation at 280 nm, were recorded at 25.0(±0.1)°C. All fluorescence measurements were performed in a thermostated Aminco-Bowman Series 2 spectrophotometer from Spectronic Instruments.

Thermal denaturation curves

The thermal denaturation curves of wild-type and mutant apoflavodoxins were followed using four spectroscopic techniques: far-UV CD, near-UV CD, near-UV absorbance and near-UV fluorescence emission. All curves were recorded, in the equipment described above, from approximately 275 K to 360 K, at a heating rate of 1.5–2 deg.C per minute. Far-UV CD curves were recorded at 222 nm and near-UV ones at 288 nm using 1 mm and 1 cm path-length cuvettes, respectively, and 30 μM protein samples in 50 mM Mops (pH 7.0). The absorbance data were obtained during the near-UV CD experiments. Fluorescence emission was recorded, for 2 μM protein samples in 50 mM Mops (pH 7.0), exciting at 280 nm and collecting the ratio of emission intensities at 320 nm and 380 nm, to minimise the dependency of fluorescence emission with temperature.

Data analyses

The urea-unfolding of apoflavodoxin is two-state.[30,33] Urea-denaturation curves were thus fitted to the following two-state equation:[47]

$$F = \frac{F_N + m_N U + (F_D + m_D U)e^{-\Delta G/RT}}{1 + e^{-\Delta G/RT}} \quad (4)$$

with:

$$\Delta G = m(U - U_m) \quad (5)$$

where F is the fluorescence signal; U is the urea concentration; U_m is the urea concentration of mid-denaturation; F_N and F_D are, respectively, fluorescence signals of the native and denatured states in the absence of urea; m_N and m_D are slopes that describe the spectroscopic signal dependence of the native and denatured states with the concentration of denaturant; and m is a linear dependence factor of free energy with urea concentration.

The thermal unfolding of apoflavodoxin is three-state.[28,48] The four thermal unfolding curves recorded for each protein using different spectroscopic techniques were thus globally fitted to the following three-state equation, derived from the two-state equation for thermal unfolding previously described:[49]

with:

$$\Delta G_1 = \Delta H_1\left(1 - \left(T/T_{m1}\right)\right) - \Delta C_{p1}\left(\left(T_{m1} - T\right) + T\log\left(T/T_{m1}\right)\right) \quad (7)$$

$$\Delta G_2 = \Delta H_2(1 - (T/T_{m2})) - \Delta C_{p2}((T_{m2} - T) + T\log(T/T_{m2})) \quad (8)$$

where Y is the observed spectroscopic signal; Y_N, Y_I and Y_D are spectroscopic signals at $T = 0$ K of native, intermediate and denatured states, respectively; m_N, m_I and m_D are the slopes of the spectroscopic signal dependence of the native, intermediate and denatured states with temperature, respectively; ΔH_1, T_{m1} and ΔC_{p1} correspond, respectively, to the enthalpy change, the melting temperature and the change in heat capacity of the first transition; and ΔH_2, T_{m2} and ΔC_{p2} to those of the second transition.

The curves corresponding to each protein were roughly normalized in order to obtain similar property value spans of approximately 1 to 0. In this way, each curve contributes similarly to the fitting. The four curves of each protein were then globally fitted to common ΔH, T_m and ΔC_p values, using the MLAB program. Since the standard deviation estimations provided by this program are not correct for non-linear models, better standard deviation values have been calculated from the 67% confidence level, as described.[50] The errors reported in this work have been calculated in that way and are usually smaller than the standard deviations given by the MLAB program.

Calculation, from urea-unfolding, of the conformational stability of the intermediate at the mid temperature of the N-to-I-equilibrium

At the temperature of mid-denaturation of the N-to-I equilibrium in the thermal unfolding of wild-type apoflavodoxin (44.2 °C), the native and intermediate states are equally populated (50% each). At that temperature, the conformational stability of the intermediate (relative to the unfolded state: ΔG_{ID}) can be determined from urea-unfolding curves, assuming that the urea unfolding of the intermediate is two-state, as it is that of the native state at 25.0 °C. Equation (9) relates the observed spectroscopic signal of the protein (Y) with those of the native, intermediate and denatured states in the absence of urea (Y_N, Y_I and Y_D, respectively); their linear dependence of urea concentration (m_n, m_i and m_d); and the free energy change, its linear dependence with urea concentration, and urea concentration of mid-denaturation of the N-to-D transition (ΔG_N, m_N and U_N) and of the I-to-D one (ΔG_I, m_I and U_I):

$$Y = \frac{Y_n + m_n U + (Y_i + m_i U)e^{-\Delta G_I/RT}e^{\Delta G_N/RT} + (Y_d + m_d U)e^{-\Delta G_I/RT}}{1 + e^{-\Delta G_I/RT}e^{\Delta G_N/RT} + e^{-\Delta G_I/RT}} \quad (9)$$

Fitting the urea-unfolding curve obtained at T_{m1} to equation (9), using for m_N its value independently calculated at 25.0 °C, allows the determination of m_I, U_N and U_I values and thus to calculate ΔG_{ID} at 44.2 °C. In order to test the simplifying assumptions implicit in equation (1) (see below), this ΔG_{ID} value has been

compared with that calculated from the global analysis of thermal unfolding curves.

Calculation of buried surface area between individual structural elements and the rest of the protein in the native structure

The SAS of each of the secondary structure elements and main loops in apoflavodoxin (and those of their corresponding complementary peptides) have been calculated in hypothetical isolation (while retaining their native conformations), using GETAREA.[51] The surface (polar or apolar) that is buried upon docking of any of the preformed structural elements onto the rest of the protein is calculated as the difference between the SAS of the two complementary peptides and that of the entire protein (PDB code 1ftg). Percentages of apolar surface burial relative to total surface burial ($100 \times SAS_{apolar}/SAS_{total}$) have been calculated for each structural element considered.

NMR spectroscopy

^1H NMR spectra were recorded for the E20K/E72K-apoflavodoxin mutant (that displays a higher accumulation of intermediate than the wild-type protein) on a Bruker DMX-600 pulse spectrometer operating at a proton frequency of 600.13 MHz. Samples for NMR experiments were prepared at a 0.5–1.0 mM concentration in $H_2O/^2H_2O$ (9 : 1, v/v) solutions containing 25 mM sodium phosphate at pH 7.0. The pH was measured with a glass microelectrode and was not corrected for isotope effects. The temperature of the NMR probe was calibrated using a methanol sample. Sodium [3-trimethylsilyl 2,2,3,3-$^2H_4$] propionate (TSP) was used as an internal reference. One-dimensional spectra were acquired using 32K data points, which were zero-filled to 64K data points before performing the Fourier transformation. Total correlated spectroscopy[52] (TOCSY) and nuclear Overhauser enhancement spectroscopy[53,54] (NOESY) spectra were recorded by standard techniques using presaturation of the water signal and the time-proportional phase incrementation mode[55]. NOESY and TOCSY mixing times were 150 ms and 80 ms, respectively. Acquisition data matrices were defined by 2018×512 points in t_2 and t_1, respectively. Data were processed using the standard XWIN-NMR Bruker program on a Silicon Graphics computer. The 2D data matrix was multiplied by a square-sine-bell window function with the corresponding shift optimised for every spectrum and zero-filled to a $2K \times 1K$ complex matrix prior to Fourier transformation. Baseline correction was applied in both dimensions.

Acknowledgements

We acknowledge financial support by grants BCM2001-252 and P120/2201. L.A.C. and M.B. were supported by FPU fellowships and J.L. by a Bask Government fellowship.

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2004.06.072

Supplementary Material for this article comprises one Table.

References

1. Baum, J., Dobson, C. M., Evans, P. A. & Hanley, C. (1989). Characterization of a partly folded protein by NMR methods: studies on the molten globule state of guinea pig alpha-lactalbumin. *Biochemistry*, **28**, 7–13.
2. Hamada, D., Hoshino, M., Kataoka, M., Fink, A. L. & Goto, Y. (1993). Intermediate conformational states of apocytochrome c. *Biochemistry*, **32**, 10351–10358.
3. Uversky, V. N. & Ptitsyn, O. B. (1994). "Partly folded" state, a new equilibrium state of protein molecules: four-state guanidinium chloride-induced unfolding of beta-lactamase at low temperature. *Biochemistry*, **33**, 2782–2791.
4. Zhu, H., Celinski, S. A., Scholtz, J. M. & Hu, J. C. (2001). An engineered leucine zipper a position mutant with an unusual three-state unfolding pathway. *Protein Sci.* **10**, 24–33.
5. Pedroso, I., Irun, M. P., Machicado, C. & Sancho, J. (2002). Four-state equilibrium unfolding of an scFv antibody fragment. *Biochemistry*, **41**, 9873–9884.
6. Hobart, S. A., Meinhold, D. W., Osuna, R. & Colon, W. (2002). From two-state to three-state: the effect of the P61A mutation on the dynamics and stability of the factor for inversion stimulation results in an altered equilibrium denaturation mechanism. *Biochemistry*, **41**, 13744–13754.
7. Rami, B. R., Krishnamoorthy, G. & Udgaonkar, J. B. (2003). Dynamics of the core tryptophan during the formation of a productive molten globule intermediate of barstar. *Biochemistry*, **42**, 7986–8000.
8. Campos, L. A. & Sancho, J. (2003). The active site of pepsin is formed in the intermediate conformation dominant at mildly acidic pH. *FEBS Letters*, **538**, 89–95.
9. Martins, S. M., Chapeaurouge, A. & Ferreira, S. T. (2003). Folding intermediates of the prion protein stabilized by hydrostatic pressure and low temperature. *J. Biol. Chem.* **278**, 50449–50455.
10. Ellis, R. J. & van der Vies, S. M. (1991). Molecular chaperones. *Annu. Rev. Biochem.* **60**, 321–347.
11. Plesofsky-Vig, N. & Brambl, R. (1995). Disruption of the gene for hsp30, an alpha-crystallin-related heat shock protein of *Neurospora crassa*, causes defects in thermotolerance. *Proc. Natl Acad. Sci. USA*, **92**, 5032–5036.
12. Parsell, D. A., Kowal, A. S., Singer, M. A. & Lindquist, S. (2002). Protein disaggregation mediated by heat-shock protein Hsp104. *Nature*, **372**, 475–478.
13. Thomas, P. J., Qu, B. H. & Pedersen, P. L. (1995). Defective protein folding as a basis of human disease. *Trends Biochem. Sci.* **20**, 456–459.
14. Prusiner, S. B. (1998). Prions. *Proc. Natl Acad. Sci. USA*, **95**, 13363–13383.
15. van der Goot, F. G., Gonzalez-Manas, J. M., Lakey, J. H. & Pattus, F. (1991). A "molten-globule" membrane-insertion intermediate of the pore-forming domain of colicin A. *Nature*, **354**, 408–410.
16. Bychkova, V. E., Berni, R., Rossi, G. L., Kutyshenko, P. & Ptitsyn, O. B. (1992). Retinol-binding protein is in the molten globule state at low pH. *Biochemistry*, **31**, 7566–7571.

17. Uversky, V. N., Narizhneva, N. V., Ivanova, T. V., Kirkitadze, M. D. & Tomashevski, A. Y. (1997). Ligand-free form of human-fetoprotein: evidence for the molten globule state. *FEBS Letters*, **410**, 280–284.
18. Cai, S. & Singh, B. R. (2001). Role of the disulfide cleavage induced molten globule state of type A botulinum neurotoxin in its endopeptidase activity. *Biochemistry*, **40**, 15327–15333.
19. Alexandrescu, A. T., Evans, P. A., Pitkeathly, M., Baum, J. & Dobson, C. M. (1993). Structure and dynamics of the acid-denatured molten globule state of alpha-lactalbumin: a two-dimensional NMR study. *Biochemistry*, **32**, 1707–1718.
20. Schulman, B. A., Kim, P. S., Dobson, C. M. & Redfield, C. (1997). A residue-specific NMR view of the non-cooperative unfolding of a molten globule. *Nature Struct. Biol.* **4**, 630–634.
21. Zhang, O., Forman-Kay, J. D., Shortle, D. & Kay, L. E. (1997). Triple-resonance NOESY-based experiments with improved spectral resolution: applications to structural characterization of unfolded, partially folded and folded proteins. *J. Biomol. NMR*, **9**, 181–200.
22. Chamberlain, A. K. & Marqusee, S. (1998). Molten globule unfolding monitored by hydrogen exchange in urea. *Biochemistry*, **37**, 1736–1742.
23. Eliezer, D., Yao, J., Dyson, J. H. & Wright, P. E. (1998). Structural and dynamic characterization of partially folded states of apomyoglobin and implications for protein folding. *Nature Struct. Biol.* **5**, 148–155.
24. Ramboarina, S. & Redfield, C. (2003). Structural characterisation of the human α-lactalbumin molten globule at high temperature. *J. Mol. Biol.* **330**, 1177–1188.
25. Fersht, A. R., Matouschek, 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.
26. Matouschek, A., Kellis, J. T., Jr, Serrano, L. & Fersht, A. R. (1989). Mapping the transition state and pathway of protein folding by protein engineering. *Nature*, **340**, 122–126.
27. Matouschek, A., Kellis, J. T., Jr, Serrano, L., Bycroft, M. & Fersht, A. R. (1990). Transient folding intermediates characterized by protein engineering. *Nature*, **346**, 440–445.
28. Irun, M. P., Garcia-Mira, M. M., Sanchez-Ruiz, J. M. & Sancho, J. (2001). Native hydrogen bonds in a molten globule: the apoflavodoxin thermal intermediate. *J. Mol. Biol.* **306**, 877–888.
29. Genzor, C. G., Perales-Alcon, A., Sancho, J. & Romero, A. (1996). Closure of a tyrosine/tryptophan aromatic gate leads to a compact fold in apoflavodoxin. *Nature Struct. Biol.* **3**, 329–332.
30. Langdon, G. M., Jimenez, M. A., Genzor, C. G., Maldonado, S., Sancho, J. & Rico, M. (2001). *Anabaena* apoflavodoxin hydrogen exchange: on the stable exchange core of the alpha/beta(21345) flavodoxin-like family. *Proteins: Struct. Funct. Genet.* **43**, 476–488.
31. Campos, L. A., Garcia-Mira, M. M., Godoy-Ruiz, R., Sanchez-Ruiz, J. M. & Sancho, J. (2004). Do proteins always benefit from a stability increase? Relevant and residual stabilization in a three-state protein by charge optimization. *J. Mol. Biol.*, in press.
32. Vuilleumier, S., Sancho, J., Loewenthal, R. & Fersht, A. R. (1993). Circular dichroism studies of barnase and its mutants: characterization of the contribution of aromatic side chains. *Biochemistry*, **32**, 10303–10313.
33. Genzor, C. G., Beldarrain, A., Gomez-Moreno, C., Lopez-Lacomba, J. L., Cortijo, M. & Sancho, J. (1996). Conformational stability of apoflavodoxin. *Protein Sci.* **5**, 1373–1388.
34. 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.
35. Sanchez, I. E. & Kiefhaber, T. (2003). Origin of unusual phi-values in protein folding: evidence against specific nucleation sites. *J. Mol. Biol.* **334**, 1077–1085.
36. Fersht, A. R. & Sato, S. (2004). Phi-value analysis and the nature of protein-folding transition states. *Proc. Natl Acad. Sci. USA*, **101**, 7976–7981.
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. Eliezer, D., Chung, J., Dyson, H. J. & Wright, P. E. (2000). Native and non-native secondary structure and dynamics in the pH 4 intermediate of apomyoglobin. *Biochemistry*, **39**, 2894–2901.
40. Englander, S. W. & Kallenbach, N. R. (1984). Hydrogen exchange and structural dynamics of proteins and nucleic acids. *Quart. Rev. Biophys.* **16**, 521–655.
41. Bai, Y., Milne, J. S., Mayne, L. & Englander, S. W. (1993). Primary structure effects on peptide group hydrogen exchange. *Proteins: Struct. Funct. Genet.* **17**, 75–86.
42. Englander, S. W., Sosnick, T. R., Englander, J. J. & Mayne, L. (1996). Mechanisms and uses of hydrogen exchange. *Curr. Opin. Struct. Biol.* **6**, 18–23.
43. Li, R. & Woodward, C. (1999). The hydrogen exchange core and protein folding. *Protein Sci.* **8**, 1571–1590.
44. Gloss, L. M. & Matthews, C. R. (1997). Urea and thermal equilibrium denaturation studies on the dimerization domain of *Escherichia coli* Trp repressor. *Biochemistry*, **36**, 5612–5623.
45. Sancho, J., Bueno, M., Campos, L. A., Fernandez-Recio, J., Irun, M. P., Lopez, J. *et al*. (2002). The relevant stability of proteins with equilibrium intermediates. *Sci. World*, **2**, 1209–1215.
46. Fillat, M. F., Borrias, W. E. & Weisbeek, P. J. (1991). Isolation and overexpression in *Escherichia coli* of the flavodoxin gene from *Anabaena* PCC 7119. *Biochem. J.* **280**, 187–191.
47. Santoro, M. M. & Bolen, D. W. (1988). Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl alpha-chymotrypsin using different denaturants. *Biochemistry*, **27**, 8063–8068.
48. Irun, M. P., Maldonado, S. & Sancho, J. (2001). Stabilization of apoflavodoxin by replacing hydrogen-bonded charged Asp or Glu residues by the neutral isosteric Asn or Gln. *Protein Eng.* **14**, 173–181.
49. Privalov, P. L. (1979). Stability of proteins: small globular proteins. *Advan. Protein Chem.* **33**, 167–241.
50. Beechem, J. M. (1992). Global analysis of biochemical and biophysical data. *Methods Enzymol.* **210**, 37–54.
51. Fraczkiewicz, R. & Braun, W. (1998). Exact and efficient analytical calculation of the accessible surface areas and their gradients for macromolecules. *J. Comput. Chem.* **19**, 319–333.

52. Rance, M. (1987). Improved techniques for homonuclear rotating frame and isotropic mixing experiments. *J. Magn. Reson.* **74**, 557–564.
53. Jeener, J., Meier, B. H., Bachmann, P. & Ernst, R. R. (1979). Investigation of exchange processes by two-dimensional NMR spectroscopy. *J. Chem. Phys.* **71**, 4546–4553.
54. Kumar, A., Ernst, R. R. & Wüthrich, K. (1980). A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton-proton cross-relaxation networks in biological macromolecules. *Biochem. Biophys. Res. Commun.* **95**, 1–6.
55. Redfield, A. G. & Kuntz, S. D. (1975). Quadrature Fourier NMR detection: Simple multiplex for dual detection. *J. Magn. Reson.* **19**, 250–254.

Edited by C. R. Matthews

(Received 1 June 2004; received in revised form 19 August 2004; accepted 26 August 2004)