Structural Analysis of an Equilibrium Folding Intermediate in the Apoflavodoxin Native Ensemble by Small-Angle X-ray Scattering

ARTICLE *in* JOURNAL OF MOLECULAR BIOLOGY · MARCH 2011

Impact Factor: 4.33 · DOI: 10.1016/j.jmb.2010.12.027 · Source: PubMed

5 AUTHORS, INCLUDING:

Rebeca Garcia-Fandino
University of Santiago de Compostela
21 PUBLICATIONS **312** CITATIONS
SEE PROFILE

Pau Bernado
French Institute of Health and Medical Rese…
79 PUBLICATIONS **2,871** CITATIONS
SEE PROFILE

doi:10.1016/j.jmb.2010.12.027 J. Mol. Biol. (2011) 406, 604–619

Contents lists available at www.sciencedirect.com

Journal of Molecular Biology

journal homepage: http://ees.elsevier.com.jmb

Structural Analysis of an Equilibrium Folding Intermediate in the Apoflavodoxin Native Ensemble by Small-Angle X-ray Scattering

Sara Ayuso-Tejedor[1,2], Rebeca García-Fandiño[3,4], Modesto Orozco[3,5], Javier Sancho[1,2,6]* and Pau Bernadó[7]*

[1]*Departamento de Bioquímica y Biología Molecular y Celular, Facultad de Ciencias, Universidad de Zaragoza, Zaragoza 50009, Spain*
[2]*Biocomputation and Physics of Complex Systems Institute, Universidad de Zaragoza, Zaragoza 50009, Spain*
[3]*Joint IRB-BSC Program in Computational Biology, Barcelona Supercomputing Center and Institute for Research in Biomedicine, Parc Científic de Barcelona, Baldiri Reixac, 10, Barcelona 08028, Spain*
[4]*Departamento de Química Fundamental, Facultad de Ciencias, Universidade da Coruña, Campus A Zapateira s/n, 15071 La Coruña, Spain*
[5]*Departament de Bioquímica, Facultat de Biologia, Av. Diagonal 647, Barcelona 08028, Spain*
[6]*Unidad Asociada BIFI-IQFR, Consejo Superior de Investigaciones Científicas, Spain*
[7]*Institute for Research in Biomedicine, Parc Científic de Barcelona, Baldiri Reixac, 10, Barcelona 08028, Spain*

Received 6 July 2010;
received in revised form
17 December 2010;
accepted 18 December 2010
Available online
7 January 2011

Edited by K. Kuwajima

Keywords:
protein folding;
folding intermediate;
small-angle X-ray scattering;
flavodoxin;
restrained molecular dynamics

Intermediate conformations are crucial to our understanding of how proteins fold into their native structures and become functional. Conventional spectroscopic measurements of thermal denaturation transitions allow the detection of equilibrium intermediates but often provide little structural detail; thus, application of more informative techniques is required. Here we used small-angle X-ray scattering (SAXS) to study the thermal denaturation of four variants of *Anabaena* PCC 7119 flavodoxin, including the wild-type apo and holo forms, and two mutants, E20K/E72K and F98N. Denaturation was monitored from changes in SAXS descriptors. Although the starting and final points of the denaturation were similar for the flavodoxin variants tested, substantial differences in the unfolding pathway were apparent between them. In agreement with calorimetric data, analysis of the SAXS data sets indicated a three-state unfolding equilibrium for wild-type apoflavodoxin, a two-state equilibrium for the F98N mutant, and increased thermostability of the E20K/E72K mutant and holoflavodoxin. Although the apoflavodoxin intermediate consistently appeared mixed with significant amounts of either native or unfolded conformations, its SAXS profile was derived from the deconvolution of the temperature-dependent SAXS data set. The apoflavodoxin thermal intermediate was structurally close to the native state but less compact, thereby indicating incipient unfolding. The residues that foster denaturation were explored by an ensemble of equilibrium ϕ-value restrained molecular dynamics. These simulations pointed to residues located in the cofactor and partner-protein recognition regions as the initial sites of

*Corresponding authors. E-mail addresses: jsancho@unizar.es; pau.bernado@irbbarcelona.org.

Abbreviations used: SAXS, small-angle X-ray scattering; MD, molecular dynamics; FMN, flavin mononucleotide; PCA, principal component analysis; MCR, multivariate curve resolution; ALS, alternating least squares; rMD, restrained MD.

denaturation and suggest a conformational adaptation as the mechanism of action in apoflavodoxin.

Introduction

A simplistic view of protein structure tends to emphasize the opposition between the native state and the denatured ensemble of unfolded conformations. While the former is amenable to structural characterization and there is a wealth of data available for thousands of native proteins, studies of the latter fall well behind. General-purpose structural models of the denatured state have been developed[1–6] and applied with the ultimate aim of describing the unfolded state.[7–10] In addition to these extreme conformations, proteins subjected to a variety of perturbations often populate alternative partly unfolded conformations,[11] some of which are close in energy to the native state and, accordingly, can be populated under native or quasi-native conditions.[12] There is increasing evidence that these "perturbed" conformations participate in protein function[13] or, in some cases, are related to the outcome of folding diseases.[14]

Partly unfolded conformations can be studied experimentally at atomic level by NMR[15–19] and at lower resolution by other techniques such as small-angle X-ray scattering (SAXS),[20–22] Förster resonance energy transfer,[23,24] and hydrodynamic measurements.[25] The protein engineering approach, particularly when coupled to equilibrium ϕ-value analysis,[26,27] can also provide valuable structural information to outline regions that are perturbed in equilibrium intermediates and to design stable models for them.[28] Furthermore, careful combination of ϕ-value analysis with the sampling capacity of molecular dynamics (MD) allows the derivation of structural information for these intermediates.[29]

SAXS is a well-established method for the structural characterization of biomolecules in solution.[30,31] Recent developments in high-brilliance synchrotron beamlines and in software for the interpretation of SAXS profiles in terms of three-dimensional models have led to a renaissance of this method in structural biology.[32,33]

SAXS has been used to follow protein folding/unfolding at equilibrium (by changing environmental conditions such as denaturant concentration[34] or temperature[35]) and in real time using rapid-mixing[36–38] or stopped-flow devices.[39,40] In these studies, several structural parameters were monitored to characterize changes in structure and compactness along the folding pathway. Traditionally, the SAXS-derived parameters used are the radius of gyration (R_g), the maximum dimension of the particle (D_{max}), the mass distance distribution function ($p(r)$), and Kratky plots.[35] However, several studies have demonstrated that more valuable information, in terms of the spectra of pure coexisting species, can be gathered when scattering profiles are deconvoluted.[34–36]

Flavodoxin has proven to be an excellent model for characterizing the structure and energetics of partly unfolded conformations in the native ensemble and for analyzing their relationship with the native state.[41] Recently, flavodoxin has also been shown to constitute a target to develop novel antimicrobials.[12,42] At moderate temperatures, most flavodoxins populate an intermediate conformation, a structural model of which has been obtained at low resolution by equilibrium ϕ-analysis.[27,43] The intermediate displays a large native-like region resembling the native conformation[27,44] and an unfolded region encompassing a cofactor binding loop and a 20-residue loop organized as a small β-sheet. This intermediate appears to be very similar to one of the partly unfolded forms of denaturant-induced flavodoxin characterized by hydrogen/deuterium exchange experiments with NMR,[45] and to a partially unfolded conformation that is dominant in some point mutants.[46]

Here we present the thermal unfolding of four variants of flavodoxin monitored by SAXS. We show that an appropriate analysis of SAXS data obtained along the unfolding transition captures the complex unfolding behavior of apoflavodoxin and provides a structural picture of the individual species involved in the process. Further structural details of this species were explored using MD with equilibrium ϕ-values as restraints. Conformational fluctuations were found in the flavin mononucleotide (FMN) and partner-protein binding regions, thus pointing to the structural determinants of the recognition mechanism in flavodoxin.

Results

Thermal unfolding of wild-type apoflavodoxin, E20K/E72K and F98N mutants, and wild-type holoflavodoxin

Anabaena flavodoxin is used as a model to study the nonnative conformations of proteins. The thermal stability of the apo form has been described in detail elsewhere.[43,47] The occurrence of an equilibrium intermediate in thermal unfolding is

evidenced by the noncoincidence of thermal unfolding curves monitored using various spectroscopic techniques. The tryptophan fluorescence emission, near-UV absorbance, far-UV circular dichroism (CD), and near-UV CD curves do not superimpose, but they can be simultaneously fitted to a three-state model (Fig. 1a). From the global fit of the three curves, the calculated melting temperatures of the N-to-I and I-to-U equilibria are 44.1 °C and 55.9 °C, respectively. At the temperature of maximal accumulation (50.3 °C), the molar fraction of the intermediate is 0.64, and it is in equilibrium with both the native state and the unfolded state. Calorimetric analysis (Fig. 1b) confirms the occurrence of the intermediate[43] and provides similar values for transition temperatures. Interestingly, the binding of the FMN cofactor to wild-type apoflavodoxin greatly stabilizes the protein, thus increasing cooperativity and making the thermal unfolding of holoflavodoxin a simple two-state equilibrium (N-to-U).[48]

A low-resolution model of the apoflavodoxin thermal intermediate has been obtained by equilibrium ϕ-analysis.[27] Ongoing efforts to obtain a high-resolution structure have focused on the design of apoflavodoxin mutants, the intermediate of which

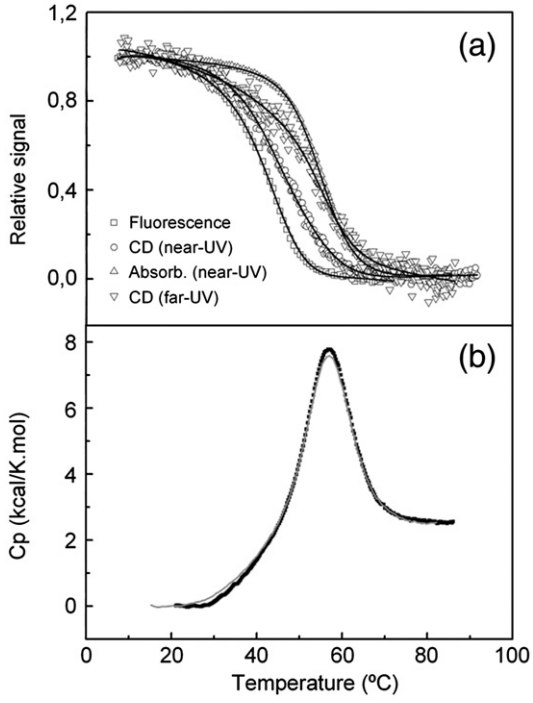

Fig. 1. (a) Thermal unfolding of wild-type apoflavodoxin, followed by fluorescence, near-UV absorbance, far-UV CD, and near-UV CD. The continuous lines represent the simultaneous fit of all data sets to a three-state equation. (b) Absolute heat capacity *versus* temperature profile for wild-type apoflavodoxin unfolding. The continuous line is the best fit to a three-state model.

populates higher molar fractions. One such mutant is E20K/E72K, which combines two previously reported mutations that stabilize the intermediate.[47] According to our analysis, the E20K/E72K intermediate is significantly stabilized compared to the unfolded state (transition temperatures of 40.6 °C and 65.7 °C), and its molar fraction is quite high (0.83 at 55.4 °C; unpublished results). A second apoflavodoxin variant (F98N) bears a destabilizing mutation at the putative interface between the native-like region and the unfolded region of the wild-type thermal intermediate. Consequently, the native state of the F98N mutant represents a stabilized version of the thermal intermediate and exhibits a two-state thermal unfolding equilibrium and a molar fraction of 1.0 at 25.0 °C.[28]

SAXS measurements of wild-type apoflavodoxin unfolding

SAXS measurements of wild-type apoflavodoxin were carried out from 6.9 °C to 64.8 °C in order to monitor changes in overall properties, such as size and compactness, along the unfolding of the protein. Kratky plots of the resulting scattering profiles were derived (Fig. 2a) to assess the degree of unfolding achieved in this series of experiments. At low temperatures, apoflavodoxin displays a prominent bell-shaped peak characteristic of globular compact proteins. As the temperature increases, the peak intensity progressively decreases, and at the highest temperature tested, the peak disappears, thus indicating that the polypeptide reaches an unstructured state.

The radius of gyration (R_g) was obtained from the small-angle region of the scattering profiles using Guinier's approach.[49] The evolution of the apoflavodoxin R_g with temperature (Fig. 2b) shows a considerable increase in protein size as the unfolding occurs from 16.7 Å at the lowest temperature to 26.8 Å at the highest temperature. The scattering profiles measured at low temperatures are in agreement with the available X-ray structure[44] (1FTG), as indicated by the excellent fit obtained with CRYSOL[50] (data not shown). Unfolding transitions were monitored following changes in R_g^2.[35] For apoflavodoxin, the jump in R_g^2 values involves a relatively broad range of temperatures (from 42.0 °C to 65 °C) encompassing the two thermal transitions previously studied by spectroscopy and calorimetry (see the text above). The absence of a sharp R_g^2 transition is consistent with the presence of additional species that differ from the native and unfolded states.

Additional structural information can be extracted from the distance distribution function ($p(r)$) obtained by the Fourier transformation of the entire scattering curve following the GNOM procedure.[51] At low temperatures, apoflavodoxin presents $p(r)$

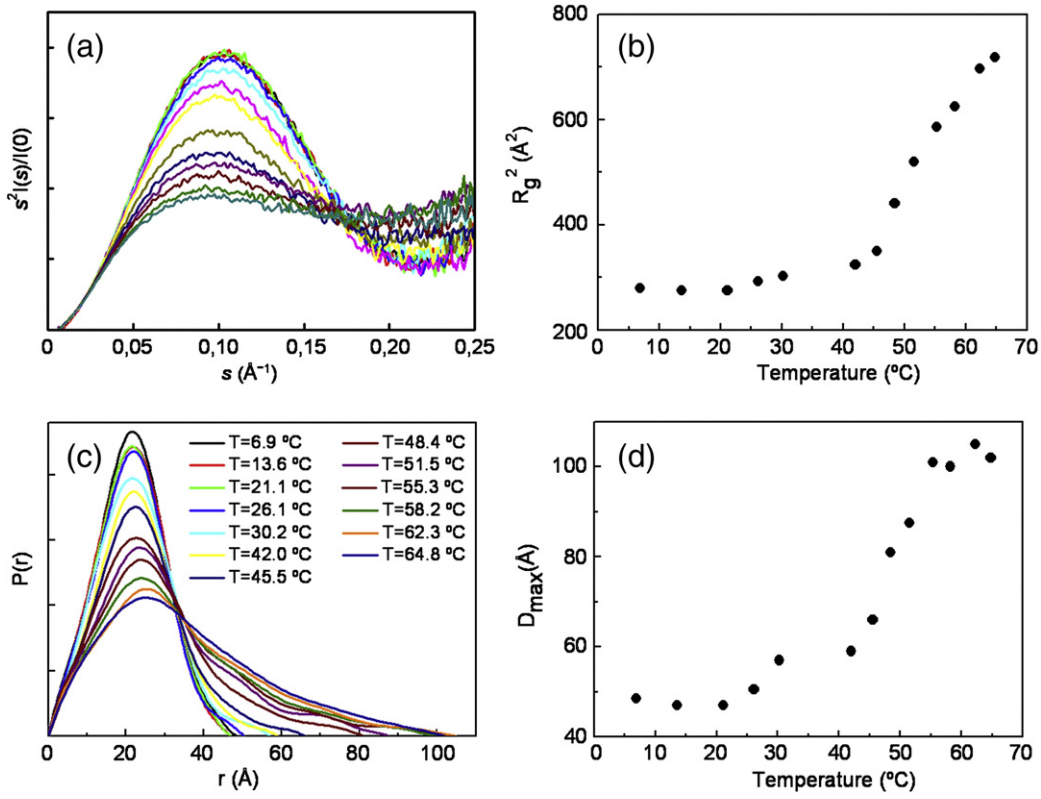

Fig. 2. Apoflavodoxin thermal unfolding followed by SAXS. (a) Kratky plots of wild-type apoflavodoxin at a range of temperatures. (b) Temperature dependence of the radius of gyration (R_g^2) of wild-type apoflavodoxin obtained by Guinier's analysis of the initial part of the curve. (c) Distance distribution function ($p(r)$) of apoflavodoxin at a range of temperatures obtained with GNOM. Specific temperatures measured are displayed in (c) and applicable to (a) and (c). (d) Temperature dependence of the maximum dimension (D_{max}) derived from the $p(r)$ function.

distributions typical of globular proteins (Fig. 2c); however, as the temperature increases, a systematic enlargement of the maximum dimension (D_{max}) of particles in solution is observed. This phenomenon smoothened the $p(r)$ function at long distances. Note that D_{max} reaches a plateau at around 102 Å at 55 °C. This finding indicates that some molecules in the ensemble achieve their maximal expansion at this temperature.

SAXS monitoring of the thermal unfolding of holoflavodoxin and point mutants E20K/E72K and F98N

To gain further insight into the conformers of apoflavodoxin that coexist along the folding equilibrium, we monitored the thermal unfolding of three additional thermodynamically well-characterized forms of this protein by SAXS. These forms were as follows: (i) wild-type holoprotein,[48] which carries one molecule of FMN; (ii) the E20K/E72K mutant, the intermediate of which is stabilized at high temperatures; and (iii) the F98N mutant, which populates the intermediate conformation at 25 °C. The Kratky plots of the three variants at a range of temperatures are shown in Fig. 3.

Holoflavodoxin and the E20K/E72K mutant display temperature-independent Kratky plots up to $T=60$ °C. After this critical point, a small increase in temperature produces dramatic decreases in peak intensity in both variants. A systematic increase in $I(s)s^2$ at large angles is also observed at higher temperatures, thereby indicating a larger population of unfolded conformations. Both holoflavodoxin and the E20K/E72K mutant display a maximum in $I(s)s^2$ function even at the highest temperature measured, thereby indicating that the thermal unfolding has not been completed. The same conclusion can be drawn from the $p(r)$ functions derived from the scattering profiles of the two variants. While the maxima of their $p(r)$ functions remained almost unchanged at around 20 Å, a slight shift towards larger radii is observed at higher temperatures.

In contrast, the F98N point mutant shows a strong temperature dependence on the scattering profiles (Fig. 3). The characteristic features of a globular particle displayed by this mutant at a low temperature (6.8 °C) are gradually transformed into those

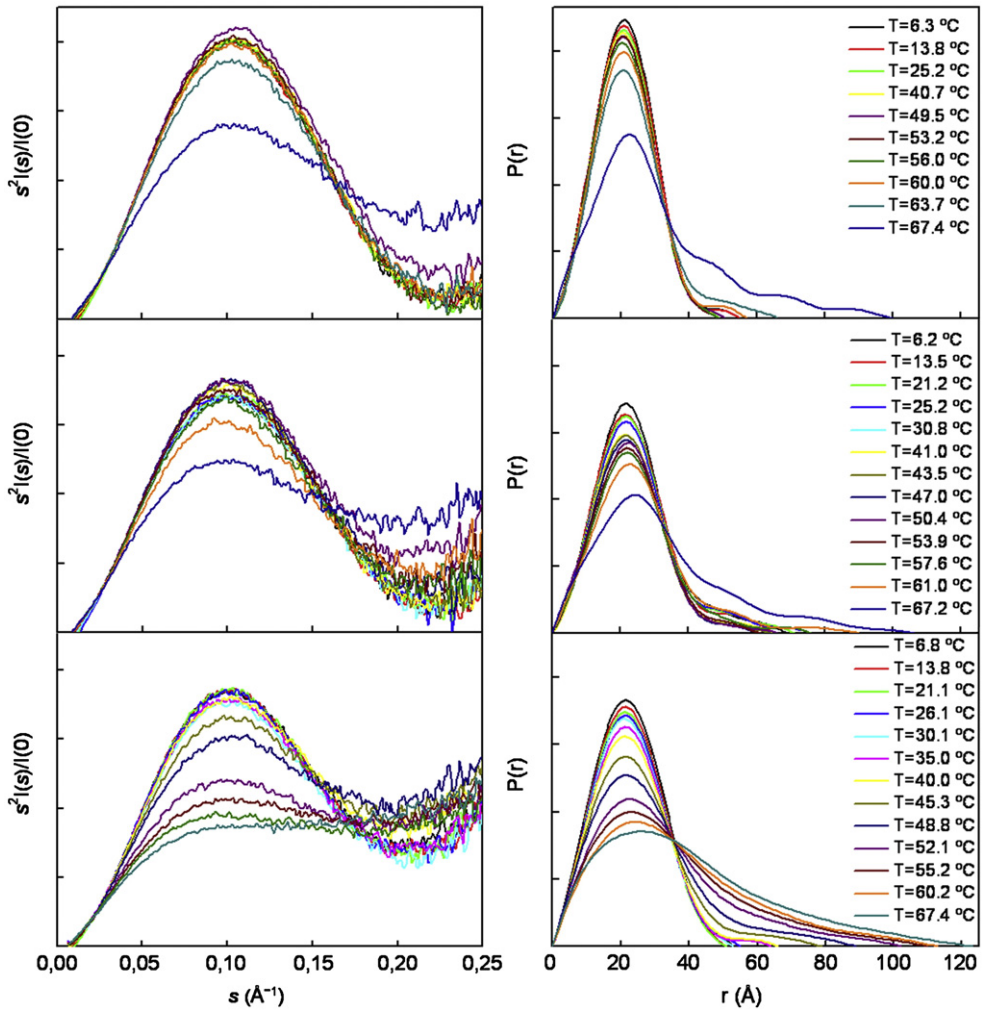

Fig. 3. Thermal unfolding of flavodoxin variants followed by SAXS. Kratky plots (left) and distance distribution functions ($p(r)$) (right) derived from the SAXS curves for the flavodoxin variants measured at a range of temperatures. Top to bottom: Wild-type holoflavodoxin, E20K/E72K, and F98N apoflavodoxin point mutants. Specific temperatures measured are displayed on the right.

of an unstructured polymer. At the highest temperature reached for this variant (67.4 °C), the Kratky plot clearly shows that the thermal unfolding has been completed and that no maximum is detected in the plot.

Figure 4 compares the evolution of R_g and D_{max} with temperature for the three flavodoxin variants with that for wild-type apoflavodoxin. At low temperatures, all the protein variants studied present R_g values (around 16 Å) very close to those expected for the X-ray structures of the apo and holo forms (15.75 Å and 15.67 Å, respectively). The change in R_g with temperature allows us to distinguish two trends within the variants. While wild-type apoflavodoxin and the F98N mutant experience a clear transition in their R_g above 45 °C, holoflavodoxin and the E20K/E72K mutant present this transition at much higher temperatures. In addition, these two latter variants attain slightly smaller R_g values, a finding that is consistent with the incomplete unfolding observed in their Kratky representations. Despite the distinct degree of unfolding reached, the four flavodoxin variants present very similar D_{max} values (around 110 Å), thereby suggesting that they share a common unfolded state.

Principal component analysis of SAXS data

Collected SAXS profiles are linear weighted averages of profiles arising from all species present in solution. Principal component analysis (PCA) of a series of experiments in which the species appear in different relative proportions can be used to determine the minimum number of species required to fully explain the data.[52] We performed a PCA (see Methods) of the four unfolding data sets corresponding to the four flavodoxin variants. The

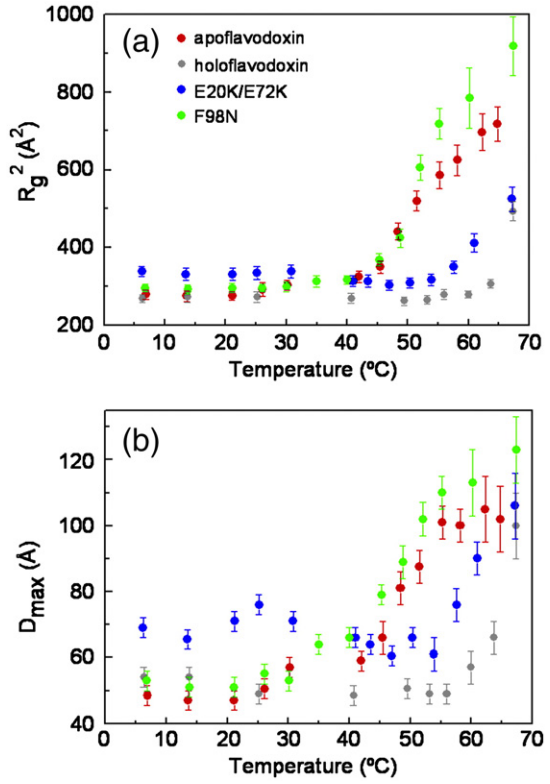

Fig. 4. Temperature evolution of (a) the square of the radius of gyration (R_g^2) and (b) the maximum dimension (D_{max}) obtained at different temperatures for the four variants: apoflavodoxin (red), holoflavodoxin (gray), E20K/E72K mutant (blue), and F98N mutant (green).

first four eigenvectors obtained for each variant are shown in Fig. 5.

Wild-type apoflavodoxin (Fig. 5a) presents three eigenvectors with features differentiated from those of the random noise observed in the fourth eigenvector. Other criteria, such as the eigenvalues and the reduction of χ^2 in data reconstructions with an increasing number of eigenvectors, also indicate the presence of three species (data not shown). Therefore, PCA substantiates the previous observation that wild-type apoflavodoxin presents an intermediate state and two thermal transitions in addition to the folded and unfolded states.[27,43]

For the E20K/E72K double mutant, our thermodynamic analysis indicates that it also populates an intermediate conformation (unpublished data). However, the relevance of the third eigenvector is less clear because of the limited range of temperatures measured (Fig. 5b). The two other variants, wild-type holoflavodoxin and the F98N mutant, show features distinguishable only from the two most important eigenvectors, and their third eigenvectors can be considered random noise (Fig. 5c and d). Therefore, the representation of the SAXS data of these variants as a combination of two species appears to be a suitable approach.

Multivariate curve resolution–alternating least squares analysis of apoflavodoxin and the F98N point mutant

The PCA described in Principal Component Analysis of SAXS Data allows us to determine the minimal number of species that have to be invoked to explain the experimental data. Unfortunately, the basis functions derived from PCA do not represent the real scattering profiles of the coexisting conformations. To obtain these profiles, we have used multivariate curve resolution (MCR)[53–57] combined with an alternating least squares (ALS) algorithm to estimate the best combination of concentration and spectral matrices (see Methods). Experimental data sets with major populations of the folded and unfolded states are required to obtain the curves of the distinct thermodynamic states.[20] The limited degree of unfolding attained for holoflavodoxin and the E20K/E72K mutant (see Fig. 4) precludes their analysis by the MCR-ALS approach. We therefore applied this analysis only to wild-type apoflavodoxin and the F98N point mutant.

The number of species required to describe the unfolding data (three and two for wild-type apoflavodoxin and the F98N mutant, respectively) were fixed in the MCR-ALS optimization. Figure 6a shows the resulting pure scattering profiles of the coexisting apoflavodoxin species as Kratky representations. As expected, the native state of apoflavodoxin presents a globular compact shape with a R_g of 16.8 ± 0.6 Å, a value in very good agreement with the X-ray structure when fitted with CRYSOL (see Fig. S1 in Supplementary Information). The D_{max} obtained from the Fourier transformation of the SAXS curve (48 ± 3 Å) is also very close to that predicted from the crystallographic structure (47 Å) (Fig. S1 in Supplementary Information). All these findings suggest that crystal structure is a reasonable model of the native conformation of the protein in solution.

The intermediate state of wild-type apoflavodoxin, obtained from MCR-ALS analysis, is notably less compact than the native state, as observed in the Kratky plot (Fig. 6a) and in the $p(r)$ function (see Fig. S1 in Supplementary Information). The decreased compactness of the intermediate is also manifested in the larger R_g (18.3 ± 0.6 Å) and D_{max} (57 ± 3 Å) values compared to those of the native state. In addition, the Kratky plot and the $p(r)$ function indicate that the intermediate is closer in structure to the native state than to the denatured ensemble. The CRYSOL fit of the intermediate scattering profile to the X-ray structure presents a reasonable agreement ($\chi = 1.17$) (see Supplementary Information) but is clearly poorer than that obtained for the native curve ($\chi = 0.55$). The structural information obtained with

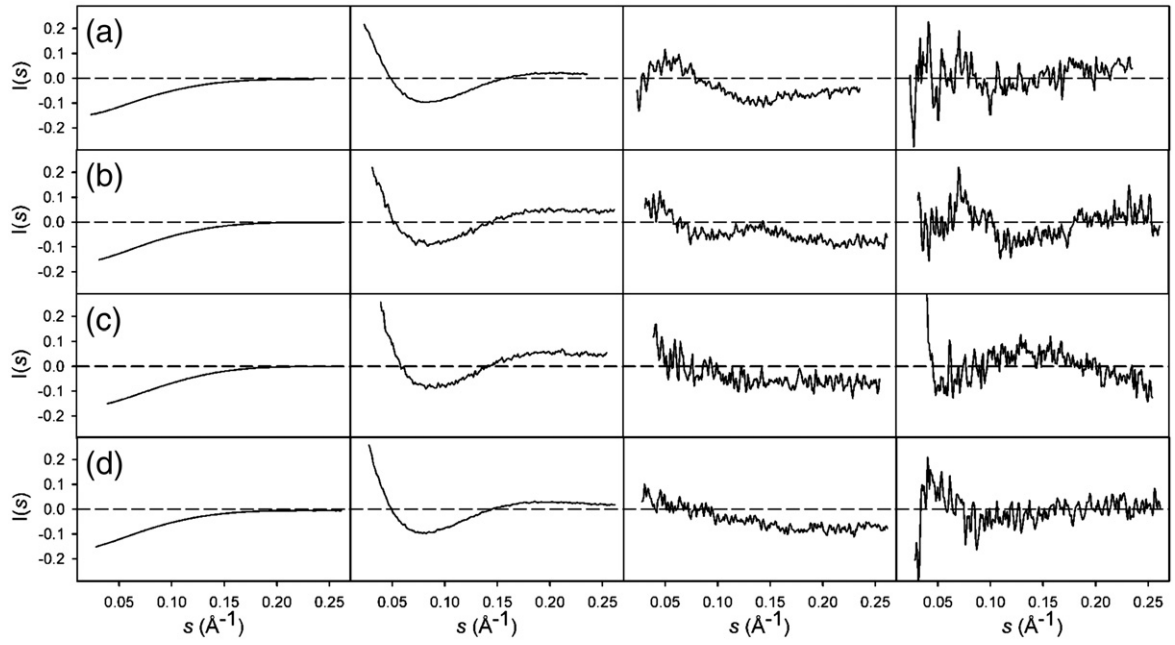

Fig. 5. First four eigenvectors (from left to right) derived from the PCA of the four flavodoxin variants studied: (a) apoflavodoxin, (b) E20K/E72K double mutant, (c) holoflavodoxin, and (d) F98N point mutant. This analysis identifies the number of species required to safely reproduce the experimental SAXS data. Different momentum transfer ranges (although similar) were used for the variants, depending on the quality of the data.

SAXS thus suggests that the thermal intermediate of wild-type apoflavodoxin is partially unfolded.

The scattering profile obtained for the more compact state of the F98N mutant presents a R_g of 17.6 ± 0.7 Å, a value larger than that of the native state and similar to that of the intermediate state of wild-type apoflavodoxin (Fig. 6b). This observation coincides with the larger D_{max} obtained for the compact state of the F98N mutant (57 ± 3 Å), which is the same value as that exhibited by the intermediate

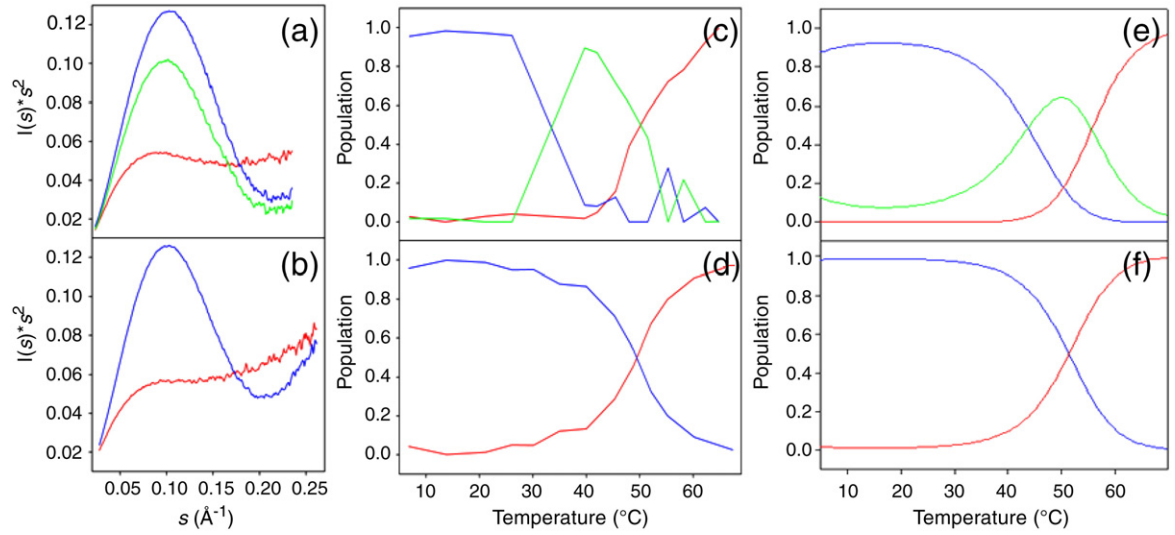

Fig. 6. Scattering profiles, in Kratky representation, derived from the unfolding series of (a) wild-type apoflavodoxin and (b) the F98N point mutant, using the MCR-ALS approach. Three states were obtained for wild-type apoflavodoxin: native (blue), intermediate (green), and unfolded (red). Two states were obtained for the F98N mutant: the more compact (blue) and the unfolded (red). Relative populations of the species obtained for (c) wild-type apoflavodoxin and (d) the F98N mutant. The same color code as in the scattering profiles was used. For comparison, the relative populations of the species derived from a spectroscopic/calorimetric analysis of (e) wild-type apoflavodoxin (from the data in Fig. 1a of this work) and (f) the F98N mutant from the data of Ayuso-Tejedor *et al.*[28] (Fig. 2a) are also shown.

of the wild-type form. Although the $p(r)$ function of the F98N compact state displays the shape expected for a globular protein, it is skewed. This observation suggests a partial degree of disorder for this species (see Fig. S1 in Supplementary Information).

The denatured states obtained from the MCR-ALS analysis of wild-type apoflavodoxin and the F98N mutant appear to be unstructured, as judged from their Kratky representations (Fig. 6). The R_g values obtained for the corresponding denatured states are 27.3 ± 1.0 Å and 28.3 ± 1.3 Å, respectively (i.e., roughly twice the values found in their native states).

In addition to the scattering profiles of the pure coexisting species, the MCR-ALS analysis of the thermal denaturation series quantifies the relative population of these species in each experiment. The MCR-ALS analysis was not linked to any thermodynamic model, and each curve was decomposed individually. Although this individualized analysis produces concentration profiles with more noise, the observation of continuous population trends for distinct species substantiates the unfolding models of the two proteins.

The SAXS-based concentration profiles for the three species present in the denaturation of wild-type apoflavodoxin are shown in Fig. 6c. A fine qualitative agreement with the more accurate concentration profiles derived from spectroscopic and calorimetric data shown in Fig. 6e is evident, providing support to the SAXS analysis protocol. From SAXS-based profiles, the temperatures of half denaturation for wild-type apoflavodoxin (N-to-I and I-to-D) are calculated to be 33.9 °C and 50.3 °C, and the corresponding enthalpy changes are

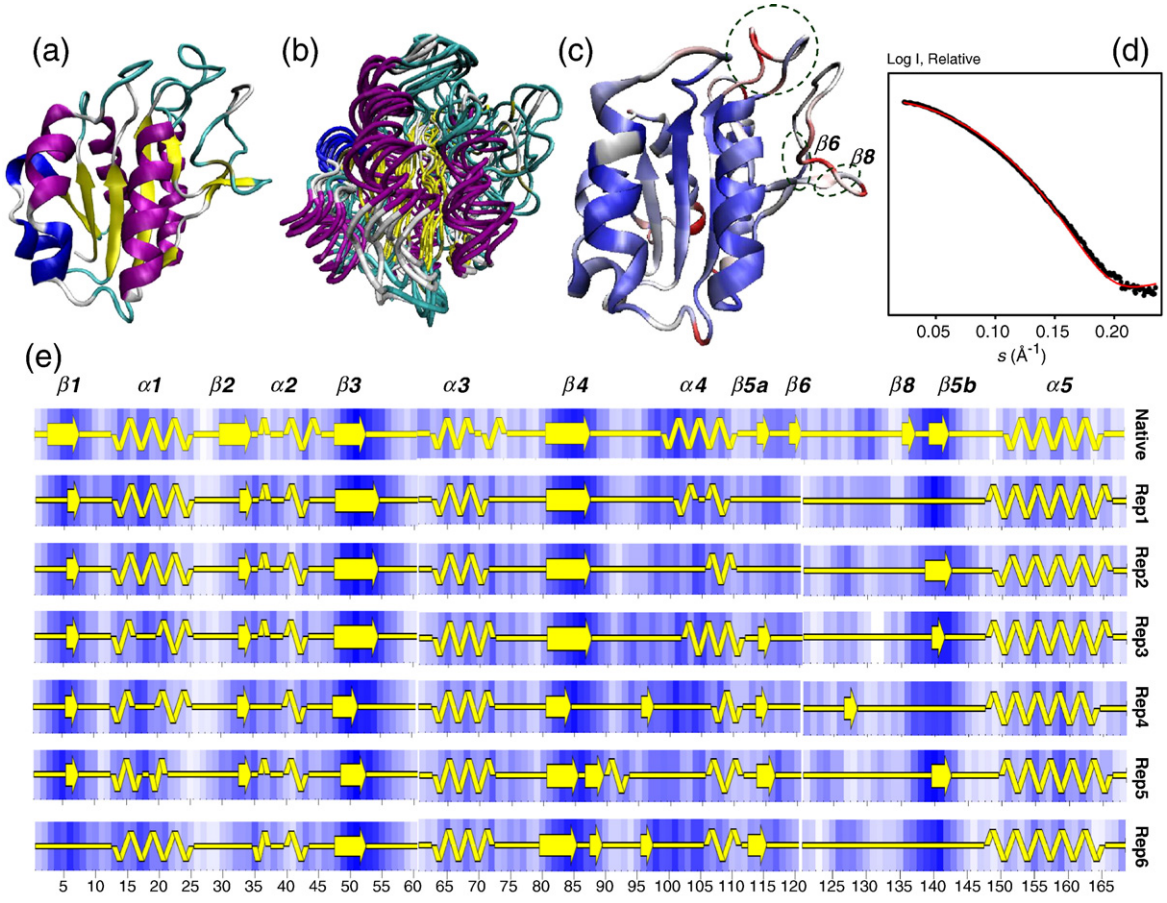

Fig. 7. Conformational variation of the ensemble obtained from the ϕ-value rMD. (a) Cartoon representation of the X-ray structure of apoflavodoxin (1FTG) colored according to the secondary structure definition (α-helix, purple; β-strand, yellow; 3_{10}-helix, blue). (b) Superimposition of the representative structures of the six clusters of the conformations found along the rMD trajectories exemplifying the conformational variation observed in the apoflavodoxin thermal intermediate; the same color code as in (a) was used. (c) Representative structure of the apoflavodoxin thermal intermediate colored according to the *B*-factor analysis at residue level along the rMD; highly flexible and static regions are displayed in red and blue, respectively. (d) Superposition of the MCR-ALS SAXS curve of the thermal intermediate species (black dots) with the theoretical SAXS curve obtained from the ensemble derived from the ϕ-value rMD simulations (red continuous line). (e) Secondary structure definition and solvent accessibility for the native conformation and the six replicas simulated for the intermediate state of apoflavodoxin following Laskowski *et al.*[64]

45.1 kcal/mol and 56.3 kcal/mol. These temperatures and enthalpy changes were reported to be 44.1 °C and 55.9 °C and 34 kcal/mol and 53 kcal/mol, respectively, from calorimetric measures (Fig. 6e).[27] While the agreement in the enthalpy changes is excellent, we found some deviations in the temperatures of half denaturation. For the first transition, this observation may be attributed to the limited number of SAXS points measured in this range. For the second transition, the lower T_m determined from the SAXS profiles may be due to differences between the true solution temperature of the SAXS experiment and the reading provided by the less sophisticated heating system used in the SAXS setup. This latter interpretation is supported by the fine correspondence offered in both T_m and enthalpy changes by the SAXS profiles (49.9 °C and 44.3 kcal/mol) and the calorimetric profiles (51.0 °C and 46.0 kcal/mol[28]) of the two-state F98N mutant (see Fig. 6d and f).

A structural model of the apoflavodoxin thermal intermediate from equilibrium φ-value restrained MD simulation

The thermodynamic properties of several point mutants of apoflavodoxin have been reported and used to derive equilibrium ϕ-values.[27,43,47,58,59] From these, a qualitative picture of the thermal intermediate ensemble can be derived by interpreting that the environment of residues with small ϕ-values is structurally perturbed in the intermediate structure with respect to the native one. A higher-resolution structure of the thermal intermediate can be derived by incorporating these data into multiple restrained MD (rMD) simulations, a strategy that has been widely used to describe the transition states of folding.[60–63]

We performed six rMD simulations of apoflavodoxin using the available set of 32 equilibrium ϕ-values as structural restraints (see Supplementary Information and Methods for details). The ensemble of structures collected in the simulation (Fig. 7b) satisfactorily reproduces the set of ϕ-values obtained experimentally. The ensemble illustrates that the general apoflavodoxin fold is maintained in the thermal intermediate because the central hydrophobic core is not perturbed. A clustering analysis of the structures collected (using a cluster radius of 5.0 Å) led to a representative structure of the thermal intermediate that can be compared with the native form found in the crystal (Fig. 7c). The secondary structure is generally preserved along the trajectories, although the termini of some α-helices and β-strands tend to unfold. Despite the preservation of the overall fold, enhanced flexibility is observed in the 90–100 region of the thermal intermediate when compared to the native structure. In addition, a conformational perturbation in β6 and β8, belonging to the so-called flavodoxin long loop,[65] is manifested in the rMD trajectories (Fig. 7e). A more detailed and quantitative picture of the structural perturbations of the intermediate with respect to the native state was derived from the comparison of the rMD simulations of the intermediate and a long (100 ns) unrestrained MD trajectory of apoflavodoxin in water. Trajectory-averaged backbone dihedral angles show that many residues placed at the two regions identified as disordered in the intermediate present values that are located in forbidden or low-populated regions of Ramachandran space (Fig. S3 in Supplementary Information). These angular violations arise from the averaging of the different conformations sampled as a result of the high level of flexibility of these regions. Importantly, these residues in the native state lie in highly probable regions, thus indicating the rigidity of this protein form. The increased flexibility observed in these two regions originates from the loss of anchoring points of these residues with the rest of the protein. This observation can be quantified through the number of native contacts present along the trajectories of the native and intermediate states when compared with the crystallographic structure (Fig. S4 in Supplementary Information). The native state maintains the contacts found in the crystal, with few exceptions (corresponding to residues in loops). Interestingly, the intermediate presents the same contacts as the native state for the first 80 residues of the protein. This observation indicates the preservation of the fold for this part of the protein. However, the 85–140 region presents a systematic decrease in native contacts in the intermediate when compared to the native state, suggesting the enhanced degree of plasticity of this region. These anchoring points lost in the intermediate were identified through the contact maps computed for representative structures of the six rMD replicas (Fig. S5 in Supplementary Information). These maps show that the breakage of the interaction between regions surrounding residues 90 and 130 is the main factor responsible for the partial unfolding of apoflavodoxin. Additional contacts around these regions are partially weakened in some of the replicas. The N-terminal part of the protein maintains the contacts found in the crystallographic structure, with the exception of the contacts of residues 9 and 65. This structural model is consistent with the low-resolution model previously proposed on the basis of a qualitative interpretation of the equilibrium ϕ-values.[27] The two disordered sections (90–100 and 120–139) are spatially close and contribute to form, in the native state, the binding site of the FMN cofactor[44,66] and the interacting region of a variety of flavodoxin partner proteins, respectively.[67,68] Therefore, these two regions involved in the biological activity of flavodoxin appear to be the least rigid parts of the

native structure and, consequently, the starting point for the unfolding process.

The ensembles of the six ϕ-values biased rMD replicas allowed us to generate a structural model of the intermediate state (see Methods). Note that these simulations display distinct degrees of expansion (see Supplementary Information), thereby indicating a limited conformational exploration achieved for individual simulations and reinforcing the need to consider the ensemble of trajectories when interpreting ϕ-values. To calculate the contribution of each individual conformation, we computed for an average scattering profile, using CRYSOL, from the structures collected for six replicas. The resulting profile presented an apparent R_g equal to 17.1 ±0.1 Å, thus indicating that the ensemble derived from rMD is more extended than the native state and slightly more compact than the experimental one. The larger degree of expansion found in the intermediate species is emphasized when comparing the available R_g values of the rMD simulation with those obtained from the unrestricted MD simulation of apoflavodoxin in water. A systematic increase in the R_g values sampled by the intermediate is found with respect to the native state of apoflavodoxin (see Fig. S2 in Supplementary Information). Therefore, these observations indicate that MD simulations provide a self-consistent picture for the structures of these two states. The structural model of the intermediate derived from rMD simulations was validated by comparing the experimentally derived and MD-derived SAXS curves, which show excellent agreement (see Fig. 7d). Altogether, with the caution implicit in these biased MD simulations, we are quite confident on the reliability of the structural model derived for the apoflavodoxin thermal intermediate.

Discussion

Full SAXS description of the thermal unfolding equilibrium for flavodoxin

The detection and description of the species involved in the unfolding equilibria of proteins and of their relative stabilities can be achieved using several types of analysis. A simple approach is monitoring the unfolding equilibria using a variety of spectroscopic techniques. If the normalized curves can be superimposed, a two-state mechanism is assumed. In contrast, if they do not superimpose, a more complex mechanism, such as a three-state mechanism, is invoked. Further analysis of data allows the determination of the population of distinct species as a function of temperature and their spectroscopic properties. When the unfolding equilibrium is probed using calorimetry, additional thermodynamic information, such as the enthalpy, entropy, and heat capacity differences between the species, can be derived. All this information is extremely valuable, but there is a lack of structural information on the key species involved in the folding/unfolding process. Complementary techniques combining low-resolution experiments with molecular simulations should be used to fill this structural gap.

Using SAXS, we followed here the unfolding of wild-type apoflavodoxin and the two mutants E20K/E72K and F98N. In addition, we monitored the thermal unfolding of the physiological complex of apoflavodoxin with its natural cofactor FMN, which we term holoflavodoxin. The agreement between the picture derived from SAXS analysis and those previously provided by the more conventional spectroscopic and calorimetric analyses is excellent. Thus, both our SAXS data and spectroscopic and calorimetric analyses[43,47] suggest a three-state unfolding mechanism for apoflavodoxin. The evolution of the molar fractions of the native, intermediate, and denatured states calculated from SAXS data (Fig. 6c) is very similar to the one described from spectroscopy and calorimetry (Fig. 6e). The transition temperatures previously reported and those derived from SAXS data are also in reasonable agreement (see the text above), and even the enthalpies of unfolding calculated from the SAXS population analysis (Fig. 6) agree well. The agreement between SAXS data and the previous conventional analysis is also excellent for holoflavodoxin. As described previously,[48] holoflavodoxin is much more stable than apoflavodoxin and, accordingly, the SAXS curves show that the unfolding had only just begun at the temperatures used in our experimental setup. The PCA of the SAXS data is consistent with the two-state behavior previously described.[48] A comparable situation applies to the unfolding of the E20K/E72K mutant for which the thermodynamic analysis indicated that the unfolding is three-state, but with the intermediate unfolding at much higher temperatures than in wild-type apoflavodoxin. Consequently, the I-to-U transition, which is the most prominent one under SAXS analysis, had only just begun in our temperature window. SAXS clearly describes the unfolding of the F98N mutant as two-state, in agreement with the findings of our spectroscopic analysis.[28] The population–temperature profiles (Fig. 6d and f), temperatures of half denaturation, and enthalpies of unfolding calculated from SAXS data are in excellent agreement with the spectroscopic analysis.

Altogether, the results summarized in the previous paragraph strongly suggest that SAXS analysis of the thermal unfolding of proteins provides precise mechanistic and thermodynamic information, such as the number of species involved and their relative populations, and also reasonable

estimates of the enthalpy changes. Furthermore, this analysis has a clear advantage over traditional techniques because it provides direct structural information for each of the species involved in the denaturation pathway.

The flavodoxin thermal intermediate according to SAXS and rMD: Functional implications

The structural and thermodynamic characterization of intermediates is a crucial step for the accurate description of folding pathways. Furthermore, these intermediates can play a fundamental role in biological function when they are energetically close to the native state (i.e., can be partially populated in biological conditions), as in the case of apoflavodoxin. Although structural studies of intermediates are complex, here we have shown that deconvolution of SAXS data measured during the unfolding process provides a direct insight into their structural features.

MCR-ALS analysis of the SAXS data set provides pure SAXS curves for the native, intermediate, and denatured species of wild-type apoflavodoxin containing structural information on their corresponding ensembles. The SAXS curve of the native state indicates that this species is a compact and globular particle. This finding is fully consistent with the crystallographic structure.[44] The denatured state, on the basis of its Kratky representation, corresponds to a highly unstructured chain with a large R_g. The equilibrium intermediate is clearly expanded relative to the native state as monitored by its R_g (increase of 1.5 Å) and particularly its D_{max}, which is 10 Å larger. This observation suggests that the intermediate state observes an incipient process of unfolding that can display a sizeable degree of conformational diversity and that is not too far from the native conformation. The moderate expansion of the equilibrium intermediate found here is consistent with a scenario in which a large part of the protein retains a native-like conformation while other regions present nonnative conformations. Unfortunately, the specific structural features of the intermediate ensemble remain unknown because of the low-resolution nature of the SAXS data.

Multiple restricted MD simulations using equilibrium ϕ-values as weak restraints were performed in order to gain a more detailed picture of the structure of the intermediate and to complement SAXS experiments.[27] These simulations, which provide structural data in agreement with SAXS experiments (Fig. 7d), identify loops implicated in FMN cofactor binding and in the binding of partner proteins as the main regions responsible for the conformational expansion of the thermal intermediate. Conversely, the rest of the secondary structure elements remain unperturbed in the thermal intermediate with respect to the native state. These findings strongly suggest that the intermediate could be considered a distorted conformation within the native ensemble[12] and that the nonnative region is limited, keeping low the entropic difference between the native state and the intermediate state.

The thermodynamic similarity between the native species and the intermediate species, together with the observation that the residues that are prone to unfold in the intermediate appear in biologically relevant regions of the protein, suggests an unexpected functional role for the intermediate state. It is now clear that the intermediate structure is accessible at biologically relevant temperatures, and that residues in the 90–100 and 120–139 loops display a larger conformational freedom than the rest of the protein. In the apoflavodoxin X-ray structures available,[44,69] the FMN cofactor binding pocket appears occluded by the interaction of the 90–100 loop with the nearby 59′ loop. This interaction is significantly debilitated in the intermediate, thereby allowing exposure of the native FMN binding pocket. Furthermore, in the intermediate, the 120–139 loop,[65] which is involved in the binding of several flavodoxin redox partner proteins,[67,68] could easily experience the conformational adaptations required to recognize this variety of proteins.[41]

Overall, we have demonstrated that, at least for a model system such as apoflavodoxin, the novel combination of complementary techniques (SAXS with ϕ-value rMD simulations) provides a unified picture of the protein equilibrium intermediates. This new perspective will contribute to our understanding of the intrinsic dynamics of the protein and the relationship of scarcely populated conformational states with their functional processes.

Methods

Wild-type and mutant flavodoxins

Mutagenesis of the *Anabaena* flavodoxin wild-type gene[70] was performed as described in the Stratagene QuikChange kit, and the mutations were identified by sequencing the entire gene. Wild-type and mutant proteins were expressed and purified, and the FMN cofactor was removed to obtain the apo forms, as described previously.[44]

Thermal denaturation

The equilibrium thermal denaturation curves of wild-type apoflavodoxin and mutant variants were followed by far-UV CD, near-UV CD, near-UV absorbance, and near-UV fluorescence emission, essentially as described previously.[43] The thermal unfolding of wild-type *Anabaena* apoflavodoxin is three-state. Thus, the unfolding curves were analyzed using the following three-state

equation, derived from the two-state equation for thermal unfolding described by Privalov:[71]

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

where ΔG_i is given by the Gibbs–Helmholz equation:

$$\Delta G_i = \Delta H_i(1 - (T/T_{mi})) - \Delta C_{pi}((T_{mi} - T) + T\ln(T/T_{m1})) \quad (2)$$

where Y is the spectroscopic signal observed; Y_N, Y_I, and Y_D are the signals of the native, intermediate, and denatured states, respectively; m_N, m_I, and m_D are slopes that describe their dependence on temperature; and ΔH_i, T_{mi}, and ΔC_{pi} correspond, respectively, to enthalpy change, melting temperature, and change in heat capacity of either the N-to-I equilibrium or the I-to-U equilibrium. The four curves acquired for each protein were globally fitted[47] to common ΔH, T_m, and ΔC_p values. To make each curve contribute similarly to the fitting, we first roughly normalized them so that they exhibited property value spans of approximately 1 to 0.

In addition, the heat capacity of apoflavodoxin was measured as a function of temperature using a differential scanning VP-DSC microcalorimeter (Microcal LLC, Northampton, MA). Protein samples and reference solutions were degassed and carefully loaded into the cells to prevent bubble formation, and the baseline of the instrument was routinely recorded before the experiments, with both cells filled with buffer. Thermal denaturation scans of apoflavodoxin were performed with 50 μM protein solutions in 50 mM 4-morpholinepropanesulfonic acid (pH 7.0) at a scanning rate of 1 °C/min from 10 °C to 100 °C. Reversibility of unfolding was checked by sample reheating after cooling inside the calorimetric cell. Differential scanning calorimetry data analysis was performed using a three-state model.

SAXS measurements

SAXS experiments were performed on the high-brilliance beamline ID02 at the European Synchrotron Radiation Facility (Grenoble, France). The samples at concentrations of 1 mg/ml were prepared in 50 mM 4-morpholinepropanesulfonic acid buffer (pH 7). A momentum transfer range of $0.07 < s < 0.31$ Å^{-1} was measured. SAXS curves were acquired at a broad range of temperature (6–67 °C). Protein samples were pushed in a capillary to the chamber where the temperature was monitored. The sample was equilibrated inside the chamber for 5 min before measurement. An equivalent protocol was applied to measure buffer profiles. Ten successive frames, each of 1 s, were acquired for both sample and buffer. Each frame was inspected, and the presence of protein damage was discarded. The scans at each temperature were averaged and subtracted from their buffer counterpart with PRIMUS[72] using standard protocols.

Forward scattering ($I(0)$) and the effective radius of gyration (R_g) were obtained from the scattering profiles using Guinier's approximation[49] and assuming that intensity can be represented as $I(s) = I(0)\exp(-(sR_g)^2/3)$ at very small angles ($s < 1.3/R_g$). The maximum particle dimension (D_{max}) and the distance distribution function ($p(r)$) were computed from all scattering patterns with the indirect transform package GNOM.[51] The agreement of SAXS profiles with three-dimensional structures was evaluated using CRYSOL.[50]

PCA was performed with the SVD algorithm implemented in MATLAB. The data matrix was decomposed in eigenvectors with their associated eigenvalues. Critical inspection of the resulting eigenvectors provides a good estimation of the number of species coexisting along the unfolding equilibrium. The additional criteria used to assess the number of independent species were the associated eigenvalues and the degree of description of the experimental data set using a reduced number of eigenvectors.

MCR-ALS analysis of SAXS data sets

MCR is a powerful tool for advanced multivariate data analysis. In the application of data measured with SAXS, MCR assumes that variance in measured data can be decomposed in the sum of individual contributions from individual components, each defined by a "pure" curve whose relative concentration profile varies under the different conditions measured. The bilinear model on which MCR is based is described by the matrix equation:

$$D = CS^T + R \quad (3)$$

where the dimensions of the matrices are $\mathbf{D}(I,J)$, $\mathbf{C}(I,N)$, $\mathbf{S^T}(N,J)$, and $\mathbf{R}(I,J)$; I is the number of rows in data matrix $\mathbf{D}$ (the number of temperatures used for the measurements); J is the number of columns in data matrix $\mathbf{D}$ (the number of momentum transfer values available); and N is the number of chemical species contributing to the spectroscopic signal. $\mathbf{C}$ is the matrix that describes the contributions of the N components in the I different temperatures of the data matrix (relative populations of the components at the different temperatures). $\mathbf{S^T}$ is the matrix that describes the instrumental responses of these N components in the J columns of the data matrix (i.e., pure curves). Due to unavoidable experimental uncertainties, this matrix decomposition is not perfect, and the differences between the measured data and their decomposition are collected in a matrix $\mathbf{R}$ of residuals.

MCR aims to find the pure SAXS curves of the coexisting species in solution, as well as the evolution of the relative concentrations of these species in response to environmental changes. In fact, according to Eq. (3), the decomposition is ambiguous, and there is an infinite number of possible and mathematically equivalent solutions of $\mathbf{C}$ and $\mathbf{S^T}$.[73] Fortunately, it is usually possible to reduce the number of possible solutions by constraining the physical nature of the solution on the basis of prior knowledge.

Here we applied the MCR method based on ALS. A detailed description of the methodology and its applications can be found elsewhere.[53–57,74] MCR-ALS was applied to wild-type apoflavodoxin and the F98N SAXS data sets using a momentum transfer range of $0.023 < s < 0.235$ Å^{-1} and $0.028 < s < 0.261$ Å^{-1}, respectively, containing a total number of 241 and 266 intensity points per temperature.

Initial estimations of the pure scattering curves best describing the SAXS data sets were obtained using the SIMPLISMA procedure, which searches for the purest curves among the experimental ones. Once identified, these serve as starting points for the MCR-ALS procedure.[56] Curves measured at 13.6 °C, 42.0 °C, and 64.8 °C were obtained for the apoflavodoxin data set, while they were obtained at 30.1 °C and 67.4 °C for the F98N mutant.

ALS optimization was performed under constraints of nonnegativity for the scattering curves and the concentration, and under constraints of unimodality (a single-profile maximum) for the optimal SAXS curves. No thermodynamic model was imposed in the relative concentration evolution with temperature.

Equilibrium φ-value rMD simulation

To determine the thermal intermediate of apoflavodoxin from *Anabaena* PCC 7119, we applied experimental equilibrium ϕ-values using rMD simulations. A total of 34 ϕ-values are available for apoflavodoxin.[27,43,47,58,59] We disregarded the ϕ-value for two mutations, D42A and S70A, because their experimental ϕ-values are larger than unity and thus are not directly interpretable.[75–80] This left us with a set of 32 experimental ϕ-values, which was used for the rMD simulations (see Supplementary Information).

Simulations were performed with the program CHARMM[81] and with an all-atom energy function (EEF1).[82] Starting from the crystallographic structure (1FTG), we first performed a short steepest-descent minimization to remove possible steric clashes. We then heated the protein to 300 K and equilibrated it for 1.0 ns. The resulting structure, which was very close to the initial conformation, was the starting structure for the rMD simulation. Six replicas of this starting configuration, all very similar, were used. Each replica was simulated for 2 ns, with the total simulation time being 12 ns. An integration step of 2 fs was used. Temperature was kept constant at 300 K using the Nose–Hoover thermostat.[83,84] The ϕ-value restraints were introduced in the simulation as a pseudo-energy term that penalizes violations of experimentally characterized ϕ-values. The equilibrium ϕ_i^{sim} value of residue i in a particular configuration was defined as the fraction of native contacts N_i made by that residue in MD with respect to those found in the crystal structure N_i^{nat} (i.e., $\phi_i^{sim} = N_i/N_i^{nat}$). To drive the simulations toward satisfying the experimental data, we used the biased MD method of Paci and Karplus[85] and Geierhaas *et al.*[86] A reaction coordinate ρ was defined as the mean squared difference between the N^e experimental ϕ-values (ϕ_i^{exp}) and the simulated ones (ϕ_i^{cal}):

$$\rho = \frac{1}{N^e}\sum_i \left(\phi_i^{cal} - \phi_i^{exp}\right)^2 \quad (4)$$

In order not to overforce the system toward the desired low ρ values, we added an energy term to the Hamiltonian only when ρ exceeded the lowest value attained up to that point in the simulation (ρ_0), defined as:

$$E = \begin{cases} \frac{\alpha}{2}(\rho - \rho_0)^2 & \rho > \rho_0 \\ 0 & \rho \leq \rho_0 \end{cases} \quad (5)$$

where the constant α controls the weight of the bias relative to the force field.[61,62]

Structures from the last 0.4 ns of the six replicas were collected and used to compute for the average SAXS curve of the thermal intermediate of apoflavodoxin for a momentum transfer range $0 < s < 0.27$ Å^{-1} with CRYSOL[50] using standard parameters.

Equilibrium MD simulation of apoflavodoxin

The crystallographic structure of apoflavodoxin (1FTG) was used as a starting point for the equilibrium MD simulation. The crystallographic ion SO_4^{2-} was conserved for the simulation. The system was immersed in a box containing 7600 water molecules, adding 24 Na^+ and 6 Cl^- to obtain neutral systems using CMIP calculations[87] to optimize the original positions of the ions. Minimization, thermalization, and 5-ns equilibration at 300 K were carried out, and a trajectory was extended for 100 ns in the isothermic isobaric ensemble ($P = 1$ atm) using periodic boundary conditions and Particle Mesh Ewald calculations.[88] A time step of 1 fs was used in conjunction with the Rattle algorithm[89] in order to maintain bonds involving hydrogen atoms at equilibrium distances. For the sake of computer efficiency, these simulations were carried out with the NAMD[90] program using the CHARMM27 force field.[91]

Acknowledgements

We thank the ID02 team at the European Synchrotron Radiation Facility for their help and Dr. Stéphanie Finet (Institut de Minéralogie et de Physique des Milieux Condensés, France) for useful discussions. Dr. Romà Tauler (Instituto de Investigaciones Químicas y Ambientales de Barcelona, Consejo Superior de Investigaciones Científicas) is acknowledged for making MCR-ALS available, and Dr. Jascha Blobel (Institut de Recerca Biomèdica de Barcelona) is acknowledged for assisting with the use of this software. Tanya Yates (Institut de Recerca Biomèdica de Barcelona) is acknowledged for reading and correcting the English manuscript. MD simulations were carried out on the MareNostrum supercomputer at the Barcelona Supercomputing Center. This work was supported by funds from the Spanish Ministry of Education (BIO2007-63458 to P.B.; BFU2007-61476/BMC to J.S.; BIO2009-10964, Combiomed RETICS Project, and Consolider E-Science Project to M.O.) and from Diputación General de Aragón, Spain (PI078/08 to J.S.). P.B. holds a Ramón y Cajal contract partially financed by the Spanish Ministerio de Educación y Ciencia. S.A.-T. was supported by an Formación de Personal Investigador fellowship from the Spanish Government. R.G.-F. was supported by the Spanish Ministerio de Educación y Ciencia and Xunta de Galicia (Programa de Becas Postdoctorales and Ángeles Alvariño, respectively).

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2010.12.027

References

1. Bernado, P., Blackledge, M. & Sancho, J. (2006). Sequence-specific solvent accessibilities of protein residues in unfolded protein ensembles. *Biophys. J.* **91**, 4536–4543.
2. Bernado, P., Blanchard, L., Timmins, P., Marion, D., Ruigrok, R. W. & Blackledge, M. (2005). A structural model for unfolded proteins from residual dipolar couplings and small-angle X-ray scattering. *Proc. Natl Acad. Sci. USA*, **102**, 17002–17007.
3. Fitzkee, N. C. & Rose, G. D. (2004). Reassessing random-coil statistics in unfolded proteins. *Proc. Natl Acad. Sci. USA*, **101**, 12497–12502.
4. Goldenberg, D. P. (2003). Computational simulation of the statistical properties of unfolded proteins. *J. Mol. Biol.* **326**, 1615–1633.
5. Jha, A. K., Colubri, A., Freed, K. F. & Sosnick, T. R. (2005). Statistical coil model of the unfolded state: resolving the reconciliation problem. *Proc. Natl Acad. Sci. USA*, **102**, 13099–13104.
6. Zhou, H. X. (2004). Polymer models of protein stability, folding, and interactions. *Biochemistry*, **43**, 2141–2154.
7. Estrada, J., Bernado, P., Blackledge, M. & Sancho, J. (2009). ProtSA: a web application for calculating sequence specific protein solvent accessibilities in the unfolded ensemble. *BMC Bioinf.* **10**, 104.
8. Meier, S., Blackledge, M. & Grzesiek, S. (2008). Conformational distributions of unfolded polypeptides from novel NMR techniques. *J. Chem. Phys.* **128**, 052204.
9. Kohn, J. E., Millett, I. S., Jacob, J., Zagrovic, B., Dillon, T. M., Cingel, N. *et al.* (2004). Random-coil behavior and the dimensions of chemically unfolded proteins. *Proc. Natl Acad. Sci. USA*, **101**, 12491–12496.
10. Mohana-Borges, R., Goto, N. K., Kroon, G. J., Dyson, H. J. & Wright, P. E. (2004). Structural characterization of unfolded states of apomyoglobin using residual dipolar couplings. *J. Mol. Biol.* **340**, 1131–1142.
11. Brockwell, D. J. & Radford, S. E. (2007). Intermediates: ubiquitous species on folding energy landscapes? *Curr. Opin. Struct. Biol.* **17**, 30–37.
12. Cremades, N., Velazquez-Campoy, A., Martinez-Julvez, M., Neira, J. L., Perez-Dorado, I., Hermoso, J. *et al.* (2009). Discovery of specific flavodoxin inhibitors as potential therapeutic agents against *Helicobacter pylori* infection. *ACS Chem. Biol.* **4**, 928–938.
13. Bahar, I., Chennubhotla, C. & Tobi, D. (2007). Intrinsic dynamics of enzymes in the unbound state and relation to allosteric regulation. *Curr. Opin. Struct. Biol.* **17**, 633–640.
14. Dobson, C. M. (2003). Protein folding and misfolding. *Nature*, **426**, 884–890.
15. Dyson, H. J. & Wright, P. E. (2004). Unfolded proteins and protein folding studied by NMR. *Chem. Rev.* **104**, 3607–3622.
16. Redfield, C. (2004). Using nuclear magnetic resonance spectroscopy to study molten globule states of proteins. *Methods*, **34**, 121–132.
17. Vallurupalli, P., Hansen, D. F., Stollar, E., Meirovitch, E. & Kay, L. E. (2007). Measurement of bond vector orientations in invisible excited states of proteins. *Proc. Natl Acad. Sci. USA*, **104**, 18473–18477.
18. Korzhnev, D. M., Salvatella, X., Vendruscolo, M., Di Nardo, A. A., Davidson, A. R., Dobson, C. M. & Kay, L. E. (2004). Low-populated folding intermediates of Fyn SH3 characterized by relaxation dispersion NMR. *Nature*, **430**, 586–590.
19. Vendruscolo, M. (2007). Determination of conformationally heterogeneous states of proteins. *Curr. Opin. Struct. Biol.* **17**, 15–20.
20. Doniach, S. (2001). Changes in biomolecular conformation seen by small angle X-ray scattering. *Chem. Rev.* **101**, 1763–1778.
21. Bernado, P., Mylonas, E., Petoukhov, M. V., Blackledge, M. & Svergun, D. I. (2007). Structural characterization of flexible proteins using small-angle X-ray scattering. *J. Am. Chem. Soc.* **129**, 5656–5664.
22. Mylonas, E., Hascher, A., Bernado, P., Blackledge, M., Mandelkow, E. & Svergun, D. I. (2008). Domain conformation of tau protein studied by solution small-angle X-ray scattering. *Biochemistry*, **47**, 10345–10353.
23. Hoffmann, A., Kane, A., Nettels, D., Hertzog, D. E., Baumgartel, P., Lengefeld, J. *et al.* (2007). Mapping protein collapse with single-molecule fluorescence and kinetic synchrotron radiation circular dichroism spectroscopy. *Proc. Natl Acad. Sci. USA*, **104**, 105–110.
24. Kuzmenkina, E. V., Heyes, C. D. & Nienhaus, G. U. (2005). Single-molecule Förster resonance energy transfer study of protein dynamics under denaturing conditions. *Proc. Natl Acad. Sci. USA*, **102**, 15471–15476.
25. Choy, W. Y., Mulder, F. A., Crowhurst, K. A., Muhandiram, D. R., Millett, I. S., Doniach, S. *et al.* (2002). Distribution of molecular size within an unfolded state ensemble using small-angle X-ray scattering and pulse field gradient NMR techniques. *J. Mol. Biol.* **316**, 101–112.
26. Yan, S., Gawlak, G., Smith, J., Silver, L., Koide, A. & Koide, S. (2004). Conformational heterogeneity of an equilibrium folding intermediate quantified and mapped by scanning mutagenesis. *J. Mol. Biol.* **338**, 811–825.
27. Campos, L. A., Bueno, M., Lopez-Llano, J., Jimenez, M. A. & Sancho, J. (2004). Structure of stable protein folding intermediates by equilibrium phi-analysis: the apoflavodoxin thermal intermediate. *J. Mol. Biol.* **344**, 239–255.
28. Ayuso-Tejedor, S., Angarica, V. E., Bueno, M., Campos, L. A., Abián, O., Bernado, P. *et al.* (2010). Design and structure of an equilibrium protein folding intermediate. A hint into dynamical regions of proteins. *J. Mol. Biol.* **400**, 922–934.
29. Gsponer, J., Hopearuoho, H., Whittaker, S. B., Spence, G. R., Moore, G. R., Paci, E. *et al.* (2006). Determination of an ensemble of structures representing the intermediate state of the bacterial immunity protein Im7. *Proc. Natl Acad. Sci. USA*, **103**, 99–104.
30. Koch, M. H., Vachette, P. & Svergun, D. I. (2003). Small-angle scattering: a view on the properties, structures and structural changes of biological macromolecules in solution. *Q. Rev. Biophys.* **36**, 147–227.

31. Svergun, D. I. & Koch, M. H. J. (2003). Small-angle scattering studies of biological macromolecules in solution. *Rep. Prog. Phys.* **66**, 1735–1782.
32. Petoukhov, M. V. & Svergun, D. I. (2007). Analysis of X-ray and neutron scattering from biomacromolecular solutions. *Curr. Opin. Struct. Biol.* **17**, 562–571.
33. Svergun, D. I. & Koch, M. H. (2002). Advances in structure analysis using small-angle scattering in solution. *Curr. Opin. Struct. Biol.* **12**, 654–660.
34. Segel, D. J., Fink, A. L., Hodgson, K. O. & Doniach, S. (1998). Protein denaturation: a small-angle X-ray scattering study of the ensemble of unfolded states of cytochrome *c*. *Biochemistry*, **37**, 12443–12451.
35. Perez, J., Vachette, P., Russo, D., Desmadril, M. & Durand, D. (2001). Heat-induced unfolding of neocarzinostatin, a small all-beta protein investigated by small-angle X-ray scattering. *J. Mol. Biol.* **308**, 721–743.
36. Akiyama, S., Takahashi, S., Kimura, T., Ishimori, K., Morishima, I., Nishikawa, Y. & Fujisawa, T. (2002). Conformational landscape of cytochrome *c* folding studied by microsecond-resolved small-angle X-ray scattering. *Proc. Natl Acad. Sci. USA*, **99**, 1329–1334.
37. Kimura, T., Uzawa, T., Ishimori, K., Morishima, I., Takahashi, S., Konno, T. *et al.* (2005). Specific collapse followed by slow hydrogen-bond formation of beta-sheet in the folding of single-chain monellin. *Proc. Natl Acad. Sci. USA*, **102**, 2748–2753.
38. Uzawa, T., Akiyama, S., Kimura, T., Takahashi, S., Ishimori, K., Morishima, I. & Fujisawa, T. (2004). Collapse and search dynamics of apomyoglobin folding revealed by submillisecond observations of alpha-helical content and compactness. *Proc. Natl Acad. Sci. USA*, **101**, 1171–1176.
39. Arai, M., Ikura, T., Semisotnov, G. V., Kihara, H., Amemiya, Y. & Kuwajima, K. (1998). Kinetic refolding of beta-lactoglobulin. Studies by synchrotron X-ray scattering, and circular dichroism, absorption and fluorescence spectroscopy. *J. Mol. Biol.* **275**, 149–162.
40. Arai, M., Ito, K., Inobe, T., Nakao, M., Maki, K., Kamagata, K. *et al.* (2002). Fast compaction of alpha-lactalbumin during folding studied by stopped-flow X-ray scattering. *J. Mol. Biol.* **321**, 121–132.
41. Sancho, J. (2006). Flavodoxins: sequence, folding, binding, function and beyond. *Cell. Mol. Life Sci.* **63**, 855–864.
42. Cremades, N., Bueno, M., Toja, M. & Sancho, J. (2005). Towards a new therapeutic target: *Helicobacter pylori* flavodoxin. *Biophys. Chem.* **115**, 267–276.
43. 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.
44. Genzor, C. G., Perales-Alcon, A., Sancho, J. & Romero, A. (1996). Closure of a tyrosine/tryptophan aromatic gate leads to a compact fold in apo flavodoxin. *Nat. Struct. Biol.* **3**, 329–332.
45. Bollen, Y. J., Kamphuis, M. B. & van Mierlo, C. P. (2006). The folding energy landscape of apoflavodoxin is rugged: hydrogen exchange reveals nonproductive misfolded intermediates. *Proc. Natl Acad. Sci. USA*, **103**, 4095–4100.
46. Bueno, M., Cremades, N., Neira, J. L. & Sancho, J. (2006). Filling small, empty protein cavities: structural and energetic consequences. *J. Mol. Biol.* **358**, 701–712.
47. 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 stabilisation in a three-state protein by charge optimisation. *J. Mol. Biol.* **344**, 223–237.
48. Campos, L. A. & Sancho, J. (2006). Native-specific stabilization of flavodoxin by the FMN cofactor: structural and thermodynamical explanation. *Proteins*, **63**, 581–594.
49. Guinier, A. (1939). La Diffraction des Rayons X aux Très Faibles Angles: Applications à l'Etude des Phénomènes Ultra-microscopiques. *Ann. Phys. (Paris)*, **12**, 161–237.
50. Svergun, D., Barberato, C. & Koch, M. H. J. (1995). CRYSOL—a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Crystallogr.* **28**, 768–773.
51. Svergun, D. I. (1992). Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.* **25**, 495–503.
52. Jolliffe, I. T. (2002). *Principal Component Analysis*, Springer edit. (Springer) Springer, New York, NY.
53. Tauler, R., Smilde, A. & Kowalski, B. (1995). Selectivity, local rank, 3-way data-analysis and ambiguity in multivariate curve resolution. *J. Chemom.* **9**, 31–58.
54. Tauler, R. (1995). Multivariate curve resolution applied to second order data. *Chemom. Intell. Lab. Syst.* **30**, 133–146.
55. de Juan, A. & Tauler, R. (2003). Chemometrics applied to unravel multicomponent processes and mixtures—revisiting latest trends in multivariate resolution. *Anal. Chim. Acta*, **500**, 195–210.
56. de Juan, A. & Tauler, R. (2006). Multivariate curve resolution (MCR) from 2000: progress in concepts and applications. *Crit. Rev. Anal. Chem.* **36**, 163–176.
57. Jaumot, J., Gargallo, R., de Juan, A. & Tauler, R. (2005). A graphical user-friendly interface for MCR-ALS: a new tool for multivariate curve resolution in MATLAB. *Chemom. Intell. Lab. Syst.* **76**, 101–110.
58. Bueno, M., Campos, L. A., Estrada, J. & Sancho, J. (2006). Energetics of aliphatic deletions in protein cores. *Protein Sci.* **15**, 1858–1872.
59. Lopez-Llano, J., Campos, L. A., Bueno, M. & Sancho, J. (2006). Equilibrium phi-analysis of a molten globule: the 1–149 apoflavodoxin fragment. *J. Mol. Biol.* **356**, 354–366.
60. Religa, T. L., Markson, J. S., Mayor, U., Freund, S. M. & Fersht, A. R. (2005). Solution structure of a protein denatured state and folding intermediate. *Nature*, **437**, 1053–1056.
61. Paci, E., Lindorff-Larsen, K., Dobson, C. M., Karplus, M. & Vendruscolo, M. (2005). Transition state contact orders correlate with protein folding rates. *J. Mol. Biol.* **352**, 495–500.
62. Salvatella, X., Dobson, C. M., Fersht, A. R. & Vendruscolo, M. (2005). Determination of the folding transition states of barnase by using PhiI-value-restrained simulations validated by double mutant PhiIJ-values. *Proc. Natl Acad. Sci. USA*, **102**, 12389–12394.
63. Varnai, P., Dobson, C. M. & Vendruscolo, M. (2008). Determination of the transition state ensemble for the folding of ubiquitin from a combination of Phi and Psi analyses. *J. Mol. Biol.* **377**, 575–588.

64. Laskowski, R. A., Rullmannn, J. A., MacArthur, M. W., Kaptein, R. & Thornton, J. M. (1996). AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR*, **8**, 477–486.
65. Lopez-Llano, J., Maldonado, S., Bueno, M., Lostao, A., Angeles-Jimenez, M., Lillo, M. P. & Sancho, J. (2004). The long and short flavodoxins: I. The role of the differentiating loop in apoflavodoxin structure and FMN binding. *J. Biol. Chem.* **279**, 47177–47183.
66. Lostao, A., Daoudi, F., Irun, M. P., Ramon, A., Fernandez-Cabrera, C., Romero, A. & Sancho, J. (2003). How FMN binds to *Anabaena* apoflavodoxin: a hydrophobic encounter at an open binding site. *J. Biol. Chem.* **278**, 24053–24061.
67. Hall, D. A., Vander Kooi, C. W., Stasik, C. N., Stevens, S. Y., Zuiderweg, E. R. & Matthews, R. G. (2001). Mapping the interactions between flavodoxin and its physiological partners flavodoxin reductase and cobalamin-dependent methionine synthase. *Proc. Natl Acad. Sci. USA*, **98**, 9521–9526.
68. Boynton, T. O., Daugherty, L. E., Dailey, T. A. & Dailey, H. A. (2009). Identification of *Escherichia coli* HemG as a novel, menadione-dependent flavodoxin with protoporphyrinogen oxidase activity. *Biochemistry*, **48**, 6705–6711.
69. Martinez-Julvez, M., Cremades, N., Bueno, M., Perez-Dorado, I., Maya, C., Cuesta-Lopez, S. *et al.* (2007). Common conformational changes in flavodoxins induced by FMN and anion binding: the structure of *Helicobacter pylori* apoflavodoxin. *Proteins*, **69**, 581–594.
70. 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.
71. Privalov, P. L. (1979). Stability of proteins: small globular proteins. *Adv. Protein Chem.* **33**, 167–241.
72. Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J. & Svergun, D. I. (2003). PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J. Appl. Crystallogr.* **36**, 1277–1282.
73. Malinowski, E. R. (2002). *Factor Analysis in Chemistry*, 3rd edit. Wiley-Interscience, New York, NY.
74. Blobel, J., Bernado, P., Svergun, D. I., Tauler, R. & Pons, M. (2009). Low-resolution structures of transient protein–protein complexes using small-angle X-ray scattering. *J. Am. Chem. Soc.* **131**, 4378–4386.
75. 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.
76. Chamberlain, A. K. & Marqusee, S. (1998). Molten globule unfolding monitored by hydrogen exchange in urea. *Biochemistry*, **37**, 1736–1742.
77. Eliezer, D., Yao, J., Dyson, H. J. & Wright, P. E. (1998). Structural and dynamic characterization of partially folded states of apomyoglobin and implications for protein folding. *Nat. Struct. Biol.* **5**, 148–155.
78. Ramboarina, S. & Redfield, C. (2003). Structural characterisation of the human alpha-lactalbumin molten globule at high temperature. *J. Mol. Biol.* **330**, 1177–1188.
79. 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. *Nat. Struct. Biol.* **4**, 630–634.
80. 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.
81. Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. & Karplus, M. (1983). CHARMM—a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **4**, 187–217.
82. Lazaridis, T. & Karplus, M. (1999). Effective energy function for proteins in solution. *Proteins*, **35**, 133–152.
83. Nose, S. (1984). A unified formulation of the constant temperature molecular-dynamics methods. *J. Chem. Phys.* **81**, 511–519.
84. Hoover, W. G. (1985). Canonical dynamics: equilibrium phase–space distributions. *Phys. Rev. A*, **31**, 1695–1697.
85. Paci, E. & Karplus, M. (1999). Forced unfolding of fibronectin type 3 modules: an analysis by biased molecular dynamics simulations. *J. Mol. Biol.* **288**, 441–459.
86. Geierhaas, C. D., Best, R. B., Paci, E., Vendruscolo, M. & Clarke, J. (2006). Structural comparison of the two alternative transition states for folding of TI I27. *Biophys. J.* **91**, 263–275.
87. Gelpi, J. L., Kalko, S. G., Barril, X., Cirera, J., de La Cruz, X., Luque, F. J. & Orozco, M. (2001). Classical molecular interaction potentials: improved setup procedure in molecular dynamics simulations of proteins. *Proteins*, **45**, 428–437.
88. Darden, T., York, D. & Pedersen, L. (1993). Particle Mesh Ewald—an $N\cdot\log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092.
89. Andersen, H. C. (1983). Rattle—a velocity version of the shake algorithm for molecular-dynamics calculations. *J. Comput. Phys.* **52**, 24–34.
90. Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E. *et al.* (2005). Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781–1802.
91. MacKerell, A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J. *et al.* (1998). All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B*, **102**, 3586–3616.