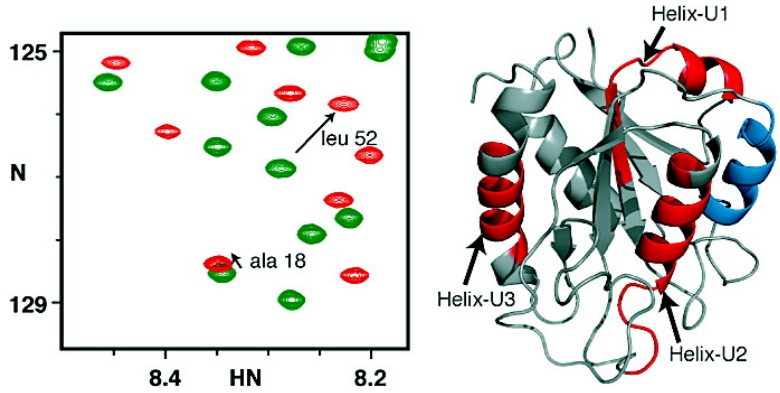
Article

Extensive Formation of Off-Pathway Species during Folding of an ### Parallel Protein Is Due to Docking of (Non)native Structure Elements in Unfolded Molecules

Sanne M. Nabuurs, Adrie H. Westphal, and Carlo P. M. van Mierlo

J. Am. Chem. Soc., **2008**, 130 (50), 16914-16920 • DOI: 10.1021/ja803841n • Publication Date (Web): 19 November 2008



More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

View the Full Text HTML

J|A|C|S
ARTICLES

Published on Web 11/19/2008

Extensive Formation of Off-Pathway Species during Folding of an α−β Parallel Protein Is Due to Docking of (Non)native Structure Elements in Unfolded Molecules

Sanne M. Nabuurs, Adrie H. Westphal, and Carlo P. M. van Mierlo*

Laboratory of Biochemistry, Wageningen University, Dreijenlaan 3, 6703 HA Wageningen, The Netherlands

Received May 22, 2008; E-mail: carlo.vanmierlo@wur.nl

Abstract: Detailed information about unfolded states is required to understand how proteins fold. Knowledge about folding intermediates formed subsequently is essential to get a grip on pathological aggregation phenomena. During folding of apoflavodoxin, which adopts the widely prevalent α−β parallel topology, most molecules fold via an off-pathway folding intermediate with helical properties. To better understand why this species is formed, guanidine hydrochloride-unfolded apoflavodoxin is characterized at the residue level using heteronuclear NMR spectroscopy. In 6.0 M denaturant, the protein behaves as a random coil. In contrast, at 3.4 M denaturant, secondary shifts and $^{1}H-^{15}N$ relaxation rates report four transiently ordered regions in unfolded apoflavodoxin. These regions have restricted flexibility on the (sub)nanosecond time scale. Secondary shifts show that three of these regions form α-helices, which are populated about 10% of the time, as confirmed by far-UV CD data. One region of unfolded apoflavodoxin adopts non-native structure. Of the α-helices observed, two are present in native apoflavodoxin as well. A substantial part of the third helix becomes β-strand while forming native protein. Chemical shift changes due to amino acid residue replacement show that the latter α-helix has hydrophobic interactions with all other ordered regions in unfolded apoflavodoxin. Remarkably, these ordered segments dock non-natively, which causes strong competition with on-pathway folding. Thus, rather than directing productive folding, conformational preorganization in the unfolded state of an α−β parallel-type protein promotes off-pathway species formation.

Introduction

Folding of proteins to conformations with proper biological activities is of vital importance to all living organisms. Current knowledge about kinetic and energetic aspects of protein folding largely stems from in vitro studies and simulations; see, for example, refs 1−3. To describe folding, the concept of a multidimensional free-energy landscape or folding funnel arose.[4] In this model, unfolded protein molecules descend along a funnel describing their free energy until they reach the state that has the lowest free energy, which is the native state. In the energy landscape model, unfolded protein molecules can fold to the native state by following different routes.

We use the expression "unfolded state" to refer to the ensemble of unfolded protein molecules at the top of the folding funnel. This state is thermodynamically separated from other folding states and corresponds to a given macroscopically observable quantity. The mean dimensions of a chemically denatured protein are those of a random coil ensemble.[5] However, it has become apparent that several unfolded proteins contain residual structure.[6,7] Simulations show that unfolded proteins with nonrandom segments can indeed exhibit random coil statistics.[8] Despite its importance, experimental information at the atomic level about structural properties of unfolded proteins is relatively sparse; see, for example, refs 6 and 9−13. This is due to conformational heterogeneity of the unfolded state and dynamic interconversion between different conformers of unfolded molecules.

Upon folding, proteins can encounter rough folding energy landscapes that allow population of partially folded species, which may be on- or off-pathway to the native state. To gain insight into events that occur early during folding, detailed information about unfolded states is crucial. Residual structure in the unfolded state is important as it most likely facilitates

(1) Dinner, A. R.; Sali, A.; Smith, L. J.; Dobson, C. M.; Karplus, M. *Trends Biochem. Sci.* **2000**, *25*, 331–339.
(2) Fersht, A. R.; Daggett, V. *Cell* **2002**, *108*, 573–582.
(3) Vendruscolo, M.; Paci, E.; Dobson, C. M.; Karplus, M. *Nature* **2001**, *409*, 641–645.
(4) Bryngelson, J. D.; Onuchic, J. N.; Socci, N. D.; Wolynes, P. G. *Proteins* **1995**, *21*, 167–195.
(5) Kohn, J. E.; Millett, I. S.; Jacob, J.; Zagrovic, B.; Dillon, T. M.; Cingel, N.; Dothager, R. S.; Seifert, S.; Thiyagarajan, P.; Sosnick, T. R.; Hasan, M. Z.; Pande, V. S.; Ruczinski, I.; Doniach, S.; Plaxco, K. W. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 12491–12496.
(6) Dyson, H. J.; Wright, P. E. *Chem. Rev.* **2004**, *104*, 3607–3622.
(7) Mittag, T.; Forman-Kay, J. D. *Curr. Opin. Struct. Biol.* **2007**, *17*, 3–14.
(8) Fitzkee, N. C.; Rose, G. D. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 12497–12502.
(9) Mok, K. H.; Kuhn, L. T.; Goez, M.; Day, I. J.; Lin, J. C.; Andersen, N. H.; Hore, P. J. *Nature* **2007**, *447*, 106–109.
(10) Neri, D.; Billeter, M.; Wider, G.; Wuthrich, K. *Science* **1992**, *257*, 1559–1563.
(11) Shortle, D.; Ackerman, M. S. *Science* **2001**, *293*, 487–489.
(12) Klein-Seetharaman, J.; Oikawa, M.; Grimshaw, S. B.; Wirmer, J.; Duchardt, E.; Ueda, T.; Imoto, T.; Smith, L. J.; Dobson, C. M.; Schwalbe, H. *Science* **2002**, *295*, 1719–1722.
(13) Religa, T. L.; Markson, J. S.; Mayor, U.; Freund, S. M.; Fersht, A. R. *Nature* **2005**, *437*, 1053–1056.

10.1021/ja803841n CCC: $40.75

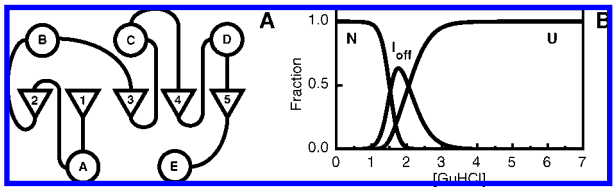


Figure 1. (a) Schematic drawing of the $\alpha-\beta$ parallel topology. α-Helices are represented as circles and β-strands as triangles. (b) Fractions of native (N), off-pathway intermediate (I_{off}), and unfolded (U) apoflavodoxin molecules as functions of GuHCl concentration.[18]

formation of folding intermediates. Knowledge about unfolded states is also relevant because some unfolded and partially folded proteins possess biological function[14] or are involved in aggregation phenomena that play a pathological role in several human diseases.[15]

Here, we investigate the guanidine hydrochloride (GuHCl)-induced unfolded state of a 179-residue flavodoxin from *Azotobacter vinelandii*. Flavodoxins consist of a single structural domain that adopts the $\alpha-\beta$ parallel topology. This topology is characterized by a parallel β-sheet flanked by α-helices (Figure 1a). Unraveling flavodoxin folding contributes to understanding how the many proteins with an $\alpha-\beta$ parallel topology fold.

Native apoflavodoxin (i.e., flavodoxin without noncovalently bound FMN cofactor) is structurally largely identical to flavodoxin[16,17] and contains no disulfide bonds. Both the GuHCl-induced equilibrium and kinetic (un)folding of flavodoxin and apoflavodoxin have been characterized.[18−20] Apoflavodoxin kinetic folding involves an energy landscape with two folding intermediates and is described by: $I_{off} \leftrightarrow$ Unfolded apoflavodoxin $\leftrightarrow I_{on} \leftrightarrow$ Native apoflavodoxin.[18] Intermediate I_{on} lies on the productive route from unfolded to native protein and is highly unstable and is thus not observed during denaturant-induced equilibrium unfolding. Approximately 90% of folding molecules fold via off-pathway intermediate I_{off}, which is a relatively stable species that needs to unfold to produce native protein and acts as a trap.[18] Apoflavodoxin equilibrium unfolding is described by: $I_{off} \leftrightarrow$ Unfolded apoflavodoxin $\leftrightarrow$ Native apoflavodoxin.[18] Intermediate I_{off} populates significantly in the concentration range of 1−3 M GuHCl (Figure 1b). The off-pathway species is compact, helical, and molten globulelike and has extremely broadened NMR resonances.[18,21] At elevated protein concentration, severe aggregation of this species occurs.[21]

Population of an off-pathway intermediate seems typical for folding of proteins with an $\alpha-\beta$ parallel topology.[22] Residual structure in the corresponding unfolded states probably facilitates formation of this species. Here, detailed information about unfolded apoflavodoxin is revealed by NMR spectroscopy. Our observations show that ordered segments exist in unfolded apoflavodoxin and that these segments dock non-natively, leading to off-pathway species formation. Apparently, proteins that contain domains with an $\alpha-\beta$ parallel topology are especially susceptible to off-pathway intermediate formation.

Materials and Methods

Sample Preparation. The single cysteine at position 69 in wild-type *A. vinelandii* (strain ATCC 478) flavodoxin II was replaced by an alanine (Cys69Ala) to avoid covalent dimerization of apoflavodoxin. Uniformly ^{15}N- and ^{13}C−^{15}N-labeled flavodoxin was obtained from transformed *Escherichia coli* cells grown on ^{15}N- and ^{13}C−^{15}N-labeled algae medium (Silantes) and purified as described.[23]

Unfolded apoflavodoxin was obtained by denaturing flavodoxin in 6 M GuHCl. Subsequently, FMN was removed via gel filtration at 7 M GuHCl. Two NMR samples of 1.5 mM unfolded ^{13}C−^{15}N-labeled Cys69Ala apoflavodoxin in 3.4 or 6.0 M GuHCl, respectively, were prepared in 10% D_2O. Samples of 0.5 mM Cys69Ala and Gln48Cys/Cys69Ala ^{15}N-labeled apoflavodoxin, both unfolded in 3.4 M GuHCl, were prepared. Gln48Cys/Cys69Ala apoflavodoxin was generated using oligo-directed mutagenesis. 2,2-Dimethyl-2-silapentane-5-sulfonic acid was present as internal chemical shift reference. DTT was present in the sample of Gln48Cys/Cys69Ala apoflavodoxin. Refractometry was used to verify the concentration of GuHCl.[24] The buffer used was 100 mM potassium pyrophosphate, pH 6.0.

Far-UV CD. CD measurements were performed on a Jasco J715 spectropolarimeter. Spectra ranging from 210 to 260 nm were recorded for 10 μM apoflavodoxin unfolded in buffer containing 3.4 or 6.0 M GuHCl, respectively, and appropriate blank spectra were subtracted. Samples were measured in a 1-mm quartz cuvette (Starna) at 25 °C. A total of 80 scans, each comprising 501 data points, were recorded per sample.

NMR Spectroscopy. Spectra were recorded on a Bruker Avance 700 MHz machine and on a Bruker DMX 500 MHz machine. Sample temperature was 25 °C.

A series of heteronuclear NMR experiments (Tables S1 and S2 in the Supporting Information) were acquired to assign backbone resonances of unfolded apoflavodoxin in 3.4 or 6.0 M GuHCl, respectively (for pulse sequences, see ref 25 and references therein).

Uniformly ^{15}N-labeled apoflavodoxin unfolded in 3.4 M GuHCl was used to determine ^{15}N longitudinal (T_1) and transverse (T_2) relaxation rates and ^{1}H−^{15}N NOE values (Table S2 in the Supporting Information).[25,26] A relaxation delay of 3 s was used during all measurements. The NOE measurements were carried out using a 3-s saturation period after a 6-s recycle delay. Experiments with and without presaturation were alternately done, and in total 12 spectra were acquired.

Data Analysis. All spectra were processed with NMRPipe[27] and analyzed using NMRViewJ.[28] The relaxation rates were determined by least squares fitting a single exponentially decaying function to the cross-peak intensity as a function of the delay in the corresponding pulse sequence. Duplicate experiments were done to

(14) Dyson, H. J.; Wright, P. E. *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 197–208.
(15) Chiti, F.; Dobson, C. M. *Annu. Rev. Biochem.* **2006**, *75*, 333–366.
(16) Steensma, E.; van Mierlo, C. P. *J. Mol. Biol.* **1998**, *282*, 653–666.
(17) Steensma, E.; Nijman, M. J.; Bollen, Y. J.; de Jager, P. A.; van den Berg, W. A.; van Dongen, W. M.; van Mierlo, C. P. *Protein Sci.* **1998**, *7*, 306–317.
(18) Bollen, Y. J.; Sanchez, I. E.; van Mierlo, C. P. *Biochemistry* **2004**, *43*, 10475–10489.
(19) Bollen, Y. J.; Nabuurs, S. M.; van Berkel, W. J.; van Mierlo, C. P. *J. Biol. Chem.* **2005**, *280*, 7836–7844.
(20) Bollen, Y. J.; Kamphuis, M. B.; van Mierlo, C. P. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 4095–4100.
(21) van Mierlo, C. P.; van den Oever, J. M.; Steensma, E. *Protein Sci.* **2000**, *9*, 145–157.
(22) Bollen, Y. J.; van Mierlo, C. P. *Biophys. Chem.* **2005**, *114*, 181–189.
(23) van Mierlo, C. P.; van Dongen, W. M.; Vergeldt, F.; van Berkel, W. J.; Steensma, E. *Protein Sci.* **1998**, *7*, 2331–2344.
(24) Nozaki, Y. The preparation of guanidine hydrochloride. In *Enzyme Structure: Part C*; Hirs, C. H. W., Timasheff, S. N., Eds.; Methods in Enzymology 26; Academic Press: New York, 1972; pp 43−50.
(25) Cavanagh, J.; Fairbrother, W. J.; Palmer, A. G.; Rance, M.; Skelton, N. J. *Protein NMR Spectroscopy: Principles and Practice*; Elsevier: Amsterdam, 2007.
(26) Farrow, N. A.; Muhandiram, R.; Singer, A. U.; Pascal, S. M.; Kay, C. M.; Gish, G.; Shoelson, S. E.; Pawson, T.; Forman-Kay, J. D.; Kay, L. E. *Biochemistry* **1994**, *33*, 5984–6003.
(27) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. *J. Biomol. NMR* **1995**, *6*, 277–293.
(28) Johnson, B. A.; Blevins, R. A. *J. Biomol. NMR* **1994**, *4*, 603–614.

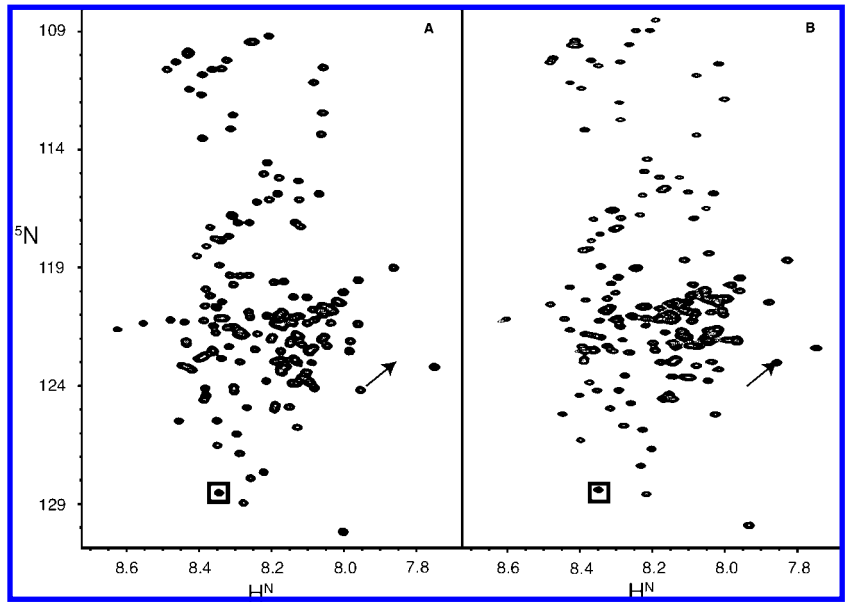


Figure 2. Gradient-enhanced 700 MHz ^{1}H−^{15}N HSQC spectra of 1.5 mM ^{13}C−^{15}N-labeled apoflavodoxin in 6.0 M (a) or 3.4 M (b) GuHCl. The box in (a) and (b) highlights an example of a cross-peak (Ala18) that does not shift upon lowering the GuHCl concentration from 6.0 to 3.4 M GuHCl. The large shift of the cross-peak of Ala169 upon lowering the denaturant concentration is shown by arrows.

estimate the standard deviations of the cross-peak intensities. NOE values were calculated from the peak intensity ratios of spectra with and without proton saturation.

The intrinsic flexibility of a residue is simulated using the radius of gyration R_g of the side chain involved.[29] The intrinsic correlation time τ for each amino acid is assumed to be proportional to R_g^3, following Stoke's Law. The decreased conformational freedom associated with proline is modeled by increasing its effective side chain R_g to 2.0 Å.[29] Relaxation rate R_2 at residue i (i.e., R_{2i}) is calculated using eq 1,[30] in which the effect of neighboring amino acids on the backbone motions of each residue is assumed to decay exponentially:[31]

$$R_{2i} = k \cdot \sum_{j=1}^{N} \tau_j \cdot \exp[-|i-j|/(\lambda_j)] \quad (1)$$

where k is an empirical scaling constant, N is the total number of residues, τ_j is the intrinsic correlation time of residue j, and λ_j is the persistence length for segmental motion of the polypeptide chain. The persistent length λ_j is set to 7 residues for all amino acid residues,[31] except for glycine and alanine for which the value of λ_j is set to 2 to account for the high flexibility of these residues.[30]

Values for the average area buried upon folding (AABUF)[32] were calculated using ProtScale.[33] A window size of 9 and a relative weight of the window edges compared to the window center of 10% were used.

Results

Upon lowering denaturant concentration, we expected the amount of residual structure in a chemically unfolded protein to increase and become better detectable. The lowest concentration of GuHCl at which unfolded apoflavodoxin is the only observable species is 3.4 M (Figure 1b). Consequently, we chose this concentration to reveal potential conformational preorganization in unfolded apoflavodoxin. To maximize random coil-like behavior of apoflavodoxin, 6.0 M GuHCl was used.

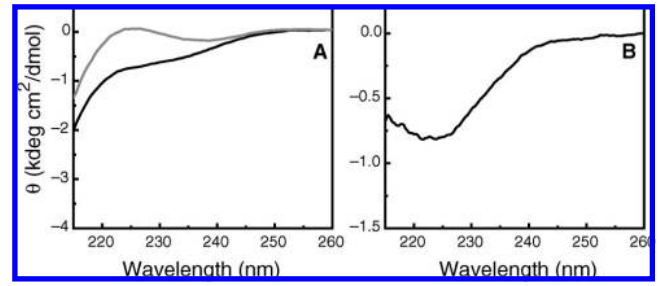


Figure 3. Far-UV CD data of unfolded apoflavodoxin. (a) The spectrum of apoflavodoxin in 6.0 M GuHCl (gray line) is typical for random coil protein. The spectrum of apoflavodoxin in 3.4 M GuHCl is shown in black. (b) Difference CD spectrum obtained by subtracting the gray spectrum from the black one in (a).

Resonance Assignments of Unfolded Apoflavodoxin. The ^{1}H−^{15}N HSQC spectra of unfolded apoflavodoxin in 3.4 and 6.0 M GuHCl are fingerprints of unfolded apoflavodoxin. Both spectra have limited ^{1}H chemical shift dispersion, which is typical for unfolded proteins (Figure 2). Assignments of all corresponding H^N, ^{15}N, ^{13}C′, $^{13}C^{\alpha}$, and $^{13}C^{\beta}$ resonances are obtained using heteronuclear NMR experiments (Tables S1 and S2 in the Supporting Information). The assignments (BioMagResBank, accession number 15474 and Tables S3 and S4 in the Supporting Information) are facilitated because of the relatively large dispersion of backbone ^{15}N and ^{13}C′ chemical shifts.

Apoflavodoxin Unfolded in 6.0 M GuHCl Behaves as a Random Coil. Far-UV CD shows that apoflavodoxin in 6.0 M GuHCl exhibits random coil behavior (Figure 3a). This behavior is confirmed by ^{1}H−^{15}N R_2 relaxation rates (Figure 4a), which are sensitive reporters of variations in polypeptide backbone dynamics.

(29) Levitt, M. *J. Mol. Biol.* **1976**, *104*, 59–107.
(30) Schwarzinger, S.; Wright, P. E.; Dyson, H. J. *Biochemistry* **2002**, *41*, 12681–12686.
(31) Schwalbe, H.; Fiebig, K. M.; Buck, M.; Jones, J. A.; Grimshaw, S. B.; Spencer, A.; Glaser, S. J.; Smith, L. J.; Dobson, C. M. *Biochemistry* **1997**, *36*, 8977–8991.
(32) Rose, G. D.; Roy, S. *Proc. Natl. Acad. Sci. U.S.A.* **1980**, *77*, 4643–4647.
(33) Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, S.; Wilkins, M. R.; Appel, R. D.; Bairoch, A. Protein Identification and Analysis Tools on the ExPASy Server. In *The Proteomics Protocols Handbook*; Walker, J. M., Ed.; Humana Press: Totowa, NJ, 2005; pp 571−607.

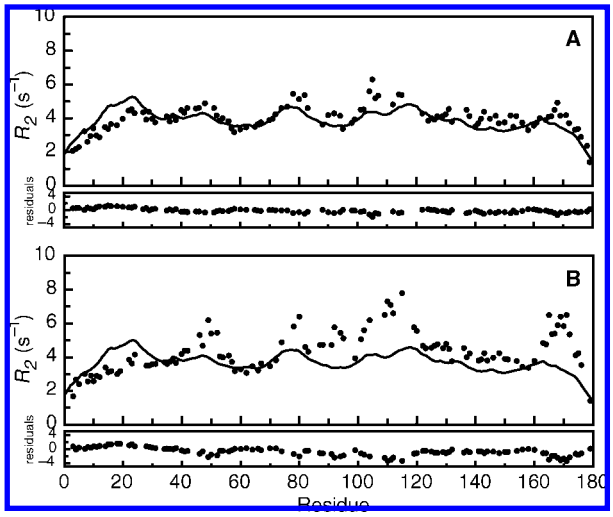


Figure 4. Sequence-dependent R_2 relaxation rates of unfolded apoflavodoxin. A model (eq 1), which takes into account sequence-dependent variations in backbone motions due to differences in radius of gyration of amino acid residues, is fitted to the R_2 values obtained at 6.0 M (a) or 3.4 M GuHCl (b). A 700 MHz spectrometer was used.

The intrinsic properties (i.e., differences in side chain size) of an amino acid sequence play a significant role in determining sequence-dependent variations in backbone dynamics of an unfolded protein.[30,31] Hence, the R_2 relaxation rates of apoflavodoxin in 6.0 M GuHCl are analyzed using a simple model that takes these variations in intrinsic flexibility of amino acid residues into account (eq 1, Materials and Methods). The model describes the R_2 relaxation rates quite well (Figure 4a). Thus, apoflavodoxin in 6.0 M GuHCl indeed behaves as a random coil and is devoid of regions with reduced flexibility.

Secondary Shifts Reveal Three Helical Regions in Unfolded Apoflavodoxin in 3.4 M GuHCl. Secondary shifts, which are deviations of observed chemical shifts from their random coil values, are good indicators for formation of secondary structure in proteins.[34,35] To properly detect possible residual structure in unfolded proteins by using secondary shifts, corrections for sequence effects should be taken into account because in particular $^{13}C'$, H^N, and ^{15}N chemical shifts are highly sensitive to local amino acid sequence.[36,37] To drastically improve the reliability for detecting residual structure within unfolded apoflavodoxin in 3.4 M GuHCl, secondary shifts are calculated using the chemical shifts of the random coil protein in 6.0 M GuHCl as intrinsic reference. Using this method, secondary shifts of amino acid residues are no longer influenced by the type of neighboring residues and thus become very sensitive reporters of secondary structure formation in unfolded apoflavodoxin.

Stretches of positive $^{13}C^{\alpha}$ and $^{13}C'$ secondary shifts reveal α-helices, whereas stretches of negative $^{13}C^{\alpha}$ and $^{13}C'$ shifts indicate β-strands. Three regions of unfolded apoflavodoxin in 3.4 M GuHCl show contiguous positive $^{13}C^{\alpha}$ and $^{13}C'$ secondary shifts (Figure 5). These α-helical regions comprise Ala41-Ile52 (Helix-U1), Glu104-Lys118 (Helix-U2), and Thr160-Ala169 (Helix-U3).

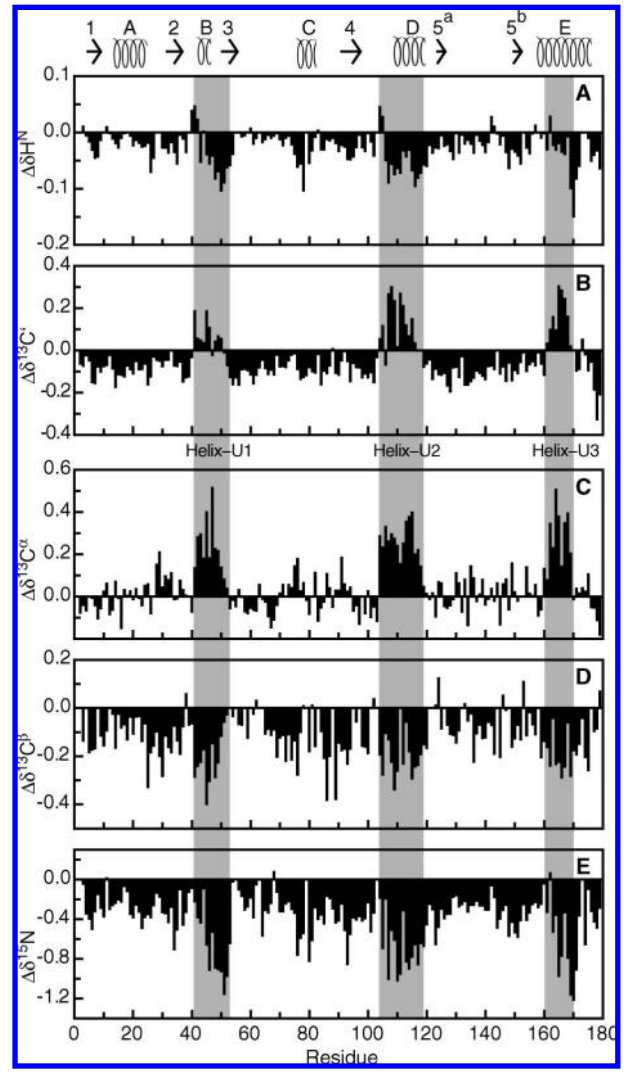


Figure 5. Secondary shifts of apoflavodoxin in 3.4 M GuHCl calculated by intrinsic referencing. Secondary shifts are shown for H^N (a), $^{13}C'$ (b), $^{13}C^{\alpha}$ (c), $^{13}C^{\beta}$ (d), and ^{15}N (e). The regions of apoflavodoxin that form α-helical structures in 3.4 M GuHCl (i.e., Helix-U1, Helix-U2, and Helix-U3) are highlighted. α-Helices and β-strands in native flavodoxin are shown at the top.

Further support for the presence of three helical regions in unfolded apoflavodoxin in 3.4 M GuHCl is obtained from H^N, $^{13}C^{\beta}$, and ^{15}N secondary shifts. Upon helix formation, these secondary shifts should be negative,[34] as is indeed observed for the majority of residues involved (Figure 5). Note that several N-terminal residues of the helices show positive H^N secondary shifts. This effect is attributed to the helix dipole,[38] which is a general property of helices and develops through alignment of amide groups with hydrogen-bonded carbonyl groups.[39] The observation that the three helices in unfolded apoflavodoxin have a dipole implies that these helices are cooperatively formed.

The transient nature of secondary structure elements in unfolded proteins causes their secondary shifts to be small. The secondary shift amplitudes of the helices detected in unfolded apoflavodoxin in 3.4 M GuHCl (Figure 5) are one-tenth of those of fully formed regular α-helices in folded proteins, which have

(34) Wishart, D. S.; Sykes, B. D. *Methods Enzymol.* **1994**, *239*, 363–392.
(35) Wang, Y.; Jardetzky, O. *Protein Sci.* **2002**, *11*, 852–861.
(36) Schwarzinger, S.; Kroon, G. J.; Foss, T. R.; Chung, J.; Wright, P. E.; Dyson, H. J. *J. Am. Chem. Soc.* **2001**, *123*, 2970–2978.
(37) Plaxco, K. W.; Morton, C. J.; Grimshaw, S. B.; Jones, J. A.; Pitkeathly, M.; Campbell, I. D.; Dobson, C. M. *J. Biomol. NMR* **1997**, *10*, 221–230.
(38) Wishart, D. S.; Sykes, B. D.; Richards, F. M. *J. Mol. Biol.* **1991**, *222*, 311–333.
(39) Hol, W. G.; van Duijnen, P. T.; Berendsen, H. J. *Nature* **1978**, *273*, 443–446.

$^{13}C^{\alpha}$ secondary shifts of about 2−3 ppm.[34] This observation implies that the helices in unfolded apoflavodoxin in 3.4 M GuHCl are populated for about 10% of the time. Chemical shifts are the only means to detect such small populations, which cannot be detected by methods based on NOEs or coupling constants.[30] Secondary shifts of $^{13}C^{\beta}$, ^{15}N, and H^N reveal no apparent secondary structure outside these helical regions.

Far-UV CD Confirms Presence of Helices in Unfolded Apoflavodoxin in 3.4 M GuHCl. The far-UV CD spectrum of unfolded apoflavodoxin in 3.4 M GuHCl suggests the presence of residual structure (Figure 3a). The difference CD spectrum, obtained by subtracting the far-UV CD spectrum of apoflavodoxin in 6.0 M GuHCl from the one obtained in 3.4 M GuHCl, has a minimum at about 222 nm (Figure 3b). This observation is typical for the presence of α-helices.

The molar ellipticity at 222 nm of native apoflavodoxin is -7.6×10^3 deg·cm^2/dmol. A fully helical 179-residue protein has an ellipticity at 222 nm of -37.5×10^3 deg·cm^2/dmol at 25 °C.[40] As the difference CD spectrum (Figure 3b) shows a molar ellipticity at 222 nm of -0.79×10^3 deg·cm^2/dmol, unfolded apoflavodoxin in 3.4 M GuHCl has a helical content of 2.1%.

NMR data show that the helical regions of unfolded apoflavodoxin in 3.4 M GuHCl comprise 37 residues, which is 21% of the total number of residues. Secondary shift amplitudes reveal that these regions populate the helical state about 10% of the time. Thus, unfolded apoflavodoxin in 3.4 M GuHCl has a helical content of 2.1%, matching the value inferred from CD data.

Relaxation Data Show That Four Regions with Restricted Flexibility Exist in Unfolded Apoflavodoxin in 3.4 M GuHCl. To further characterize unfolded apoflavodoxin in 3.4 M GuHCl, its dynamic features are probed by $^1H-^{15}N$ R_1, R_2, and NOE relaxation experiments, which reveal backbone flexibilities and conformational exchange processes on (sub)nano- to millisecond time scales.

The $^1H-^{15}N$ NOE values of unfolded apoflavodoxin are distributed around 0.23 (Figure S1a in the Supporting Information). The R_1 relaxation rates of apoflavodoxin in 3.4 M GuHCl are uniformly distributed around an average of 1.5 s^{-1} (Figure S1b in the Supporting Information). Similar average R_1 values are obtained for other unfolded proteins.[30,31] Both N- and C-terminal residues have even lower R_1 relaxation rates and negative NOE values as these residues are very flexible.

The R_2 relaxation rates of apoflavodoxin in 3.4 M GuHCl vary between about 2 and 8 s^{-1} (Figure 4b) and are much less uniformly distributed than R_1 relaxation rates. Four stretches of residues have elevated R_2 values: Ala41-Gly53, Glu72-Gly83, Gln99-Ala122, and Thr160-Gly176, implying restricted flexibility on the (sub)nanosecond time scale in these regions of the unfolded protein.

The increased R_2 relaxation rates are not due to exchange between different conformers on micro- to millisecond time scales because similar rates are observed using 500 and 700 MHz spectrometers (Figure S1c in the Supporting Information). If exchange occurs, its contribution to R_2 is proportional to the square of magnetic field strength involved. Without exchange, R_2 is only marginally affected by changes in field strength.[41]

In summary, four regions of apoflavodoxin in 3.4 M GuHCl exhibit reduced backbone flexibility on the (sub)nanosecond time scale.

Helix Formation and Hydrophobic Interactions Cause Reduced Flexibility in Unfolded Apoflavodoxin. Because four regions of unfolded apoflavodoxin have reduced flexibility on the (sub)nanosecond time scale, the simple model that takes into account sequence-dependent variations in backbone motions of an unfolded protein cannot describe the experimental R_2 data. Clearly, the elevated R_2 values of residues 41−53, 72−83, 99−122, and 160−176 are not fitted well using eq 1 (Figure 4b).

In the regions with restricted flexibility, interactions within and/or between clusters of residues most likely influence backbone motions of unfolded apoflavodoxin. Such types of interactions are also observed for urea-denatured apomyoglobin[30] and lysozyme.[12] The restricted flexibility is mainly due to helix formation in three of the four regions mentioned, as shown by analysis of secondary shifts of unfolded apoflavodoxin. The region with reduced flexibility that comprises residues 72−83 does not form α-helical secondary structure, as no positive secondary shifts of the corresponding $^{13}C^{\alpha}$ and $^{13}C'$ nuclei are observed (Figure 5). However, this region has negative H^N secondary shifts. Thus, residues 72−83 probably form a transiently ordered structure that is neither α-helix nor β-strand.

The AABUF is correlated with hydrophobicity[32] and corresponds to sequence-dependent dynamic variations due to hydrophobic interactions in unfolded proteins.[30,42] The four regions with restricted flexibility in unfolded apoflavodoxin match with regions of large AABUF (Figure 6a,b). Hydrophobic interactions will stabilize initially formed structures within unfolded apoflavodoxin.

Transient Non-Native Docking of Ordered Segments in Unfolded Apoflavodoxin. To probe via chemical shift changes whether the four ordered regions in unfolded apoflavodoxin in 3.4 M GuHCl interact with one another, use is made of site-directed mutagenesis to slightly change the hydrophobicity of one of the ordered regions. We replaced glutamine at position 48 by cysteine, which has hydrophobic characteristics at pH 6. In addition, this replacement enables site-directed spin labeling of the protein, allowing future paramagnetic relaxation enhancement experiments. Residue 48 resides in the middle of Helix-U1 in unfolded apoflavodoxin. Figure 6c shows that chemical shifts of many residues in unfolded apoflavodoxin change because this cysteine is introduced. Interestingly, several residues of Helix-U2 and Helix-U3 are also affected. Furthermore, Leu78, which resides in the region with reduced flexibility comprising residues 72−83, shows chemical shift changes. Interaction of Helix-U1 with Leu78 explains the reduced flexibility and the contiguous stretch of negative H^N secondary shifts for residues 75−81 (Figure 5a). The signs of the H^N and ^{15}N chemical shift changes upon replacing Gln48 by Cys48 reveal that these changes are due to increased population of all three helices observed in unfolded apoflavodoxin (Figure S2 in the Supporting Information). Stabilization of Helix-U2 and Helix-U3 can only occur if long-range hydrophobic interactions between these helices and Helix-U1 exist in unfolded apoflavodoxin. The chemical shift changes show that Helix-U1 has

(40) Scholtz, J. M.; Qian, H.; York, E. J.; Stewart, J. M.; Baldwin, R. L. *Biopolymers* **1991**, *31*, 1463–1470.
(41) Carrington, A.; MacLachlan, A. D. *Introduction to Magnetic Resonance*; Harper: New York, 1969.
(42) Le Duff, C. S.; Whittaker, S. B.; Radford, S. E.; Moore, G. R. *J. Mol. Biol.* **2006**, *364*, 824–835.

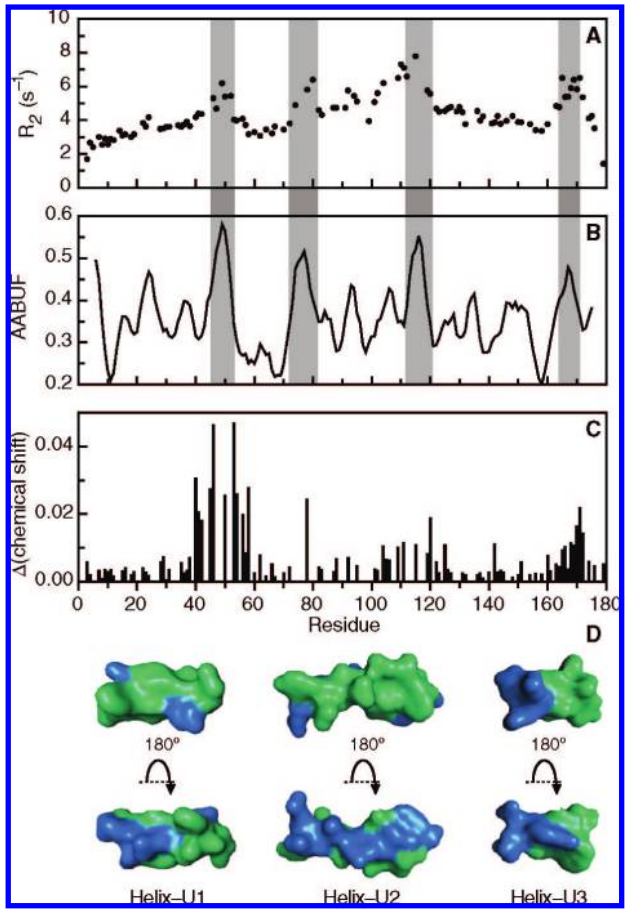


Figure 6. Four regions with elevated R_2 relaxation rates in unfolded apoflavodoxin in 3.4 M GuHCl correspond to regions with large surface area buried upon folding. (a) R_2 values of apoflavodoxin unfolded in 3.4 M GuHCl. (b) AABUF. Gray boxes highlight maxima in AABUF. (c) Differences between backbone amide chemical shifts of 48Q- and 48C-apoflavodoxin variants unfolded in 3.4 M GuHCl. Chemical shift differences are calculated using $\Delta(\text{chemical shift}) = \Delta\delta H^N + 0.1 \cdot |\Delta\delta^{15}N|$. (d) Cartoon drawing of the helical regions in unfolded apoflavodoxin. Hydrophobic residues are shown in green.

transient hydrophobic interactions with all other ordered regions in the unfolded protein and implies non-native docking of these regions.

Discussion

The most interesting and physiologically relevant situation to characterize unfolded apoflavodoxin would be under native conditions. Unfortunately, the overwhelming majority of protein molecules is then natively folded (Figure 1b). Detailed characterization of unfolded apoflavodoxin by NMR spectroscopy thus occurred in the presence of a chemical denaturant. The corresponding unfolded state may therefore differ from the unfolded state under native conditions.

Four regions of unfolded apoflavodoxin in 3.4 M GuHCl have reduced flexibility, which is due to transient helix formation and local and nonlocal hydrophobic interactions (Figure 7a,b). Residues 41−45 of Helix-U1, residues 108−118 of Helix-U2, and all residues of Helix-U3 are helical in the native protein as well. These stretches of residues thus have intrinsic propensity to form nativelike helices. Helices are the only regular secondary structure elements in unfolded apoflavodoxin in 3.4 M GuHCl, and these helices are sufficiently stable to be present about 10% of the time.

Non-native Structure Elements in Unfolded Apoflavodoxin. Remarkably, the C-terminal part (residues 47−52) of Helix-U1 in unfolded apoflavodoxin is non-native. In native protein, several of these residues form the N-terminal part of β-strand-3, which is the central strand of apoflavodoxin's parallel β-sheet (Figure 7a). The N-terminal part (residues 104−107) of Helix-U2 in unfolded apoflavodoxin also is non-native, as it is not helical in native protein. Residues 72−83 (light blue in Figure 7a) form a transiently ordered structure in unfolded apoflavodoxin that is neither α-helix nor β-strand and is involved in non-native hydrophobic interactions with Helix-U1.

Residues Ala41-Ile52 (i.e., Helix-U1) form a transient helix in unfolded apoflavodoxin. However, part of this helix (residues 49−52) becomes β-strand-3 in native protein. To further investigate the intrinsic propensity for secondary structure-type formation of this region, use is made of Scratch Protein Predictor (SSPro),[43] Jnet,[44] and AGADIR.[45] These programs use structure prediction algorithms that are based solely on propensity of individual residues to be part of regular secondary structure types. Although the following is no direct proof, these algorithms predict helix formation in a peptide fragment comprising residues Val36-Glu61 of apoflavodoxin. The helix begins at either Arg38 (AGADIR), Val39 (Jnet), or Ala41 (SSPro) and continues until Ile51 (AGADIR, Jnet, SSPro). AGADIR assigns positive $^{13}C^{\alpha}$ secondary shifts to residues Glu42-Leu52 (results not shown), matching our experimental data (Figure 5c). Formation of Helix-U1 in unfolded apoflavodoxin is an intrinsic property of its amino acid sequence. Helix-U1 does not need to be induced through contacts with other regions of the unfolded protein. These observations strongly suggest that formation of native β-strand-3 during apoflavodoxin folding requires specific tertiary contacts.

Structure Formation in Unfolded Apoflavodoxin Promotes Off-Pathway Intermediate Production. The chemical shift changes upon replacing Gln48 by Cys48 show that Helix-U1 has hydrophobic interactions with all other ordered regions in unfolded apoflavodoxin and imply transient non-native docking of these regions.

Support for conformational preorganization in unfolded apoflavodoxin promoting off-pathway intermediate formation is provided by hydrogen/deuterium (H/D) exchange results. Detection by NMR spectroscopy of native state H/D exchange in the presence of small amounts of a denaturant identified five unfolding clusters of residues within native apoflavodoxin.[20] Four of these clusters unfold subglobally in a cooperative manner. The resulting conformations are partially unfolded forms (PUFs) of the protein. Both PUF1 and PUF2 are unfolding excursions that start from native apoflavodoxin but do not continue to the unfolded state and do not reside on the productive folding route. In contrast, both PUF3 and PUF4 (Figure 7c) almost certainly are PUFs of the off-pathway folding intermediate[20] and are positioned between this species and apoflavodoxin's unfolded state. Remarkably, the study presented here reveals that the residues that are folded and protected against H/D exchange in PUF3 are characterized by enhanced R_2 relaxation rates and by formation of transient secondary structure within apoflavodoxin in 3.4 M GuHCl (Figure 7b).

(43) Cheng, J.; Randall, A. Z.; Sweredoski, M. J.; Baldi, P. *Nucleic Acids Res.* **2005**, *33*, W72–W76.
(44) Cuff, J. A.; Barton, G. J. *Proteins: Struct., Funct., Genet.* **1999**, *34*, 508–519.
(45) Munoz, V.; Serrano, L. *Nat. Struct. Biol.* **1994**, *1*, 399–409.

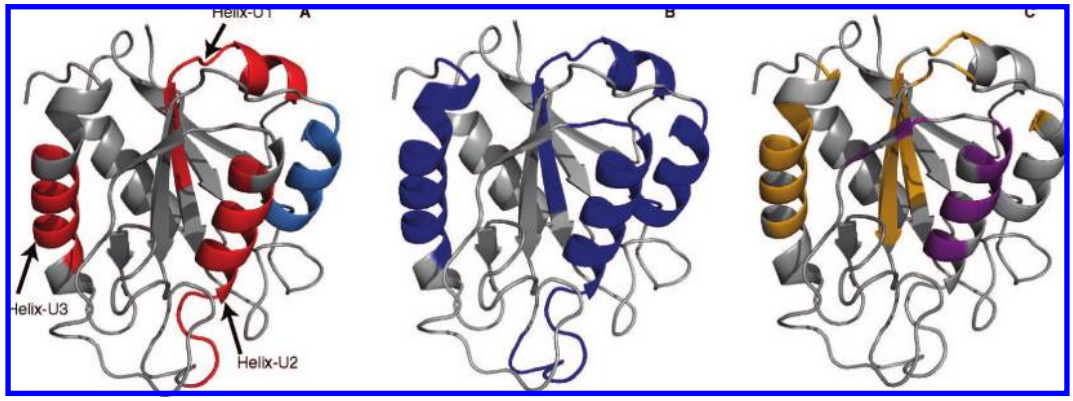


Figure 7. Cartoon drawings of flavodoxin from *A. vinelandii* (pdb ID 1YOB). The FMN cofactor is not shown. (a) Regions of unfolded apoflavodoxin in 3.4 M GuHCl that adopt α-helical structure are red. Part of β-strand-3 of native flavodoxin has helical characteristics in the unfolded protein (i.e., Helix-U1). The region that forms a transiently ordered structure that is neither α-helix nor β-strand and has non-native hydrophobic interactions with Helix-U1 is light blue. (b) Regions of unfolded apoflavodoxin in 3.4 M GuHCl with elevated R_2 relaxation rates are blue. (c) Partially unfolded forms PUF3 and PUF4 of apoflavodoxin detected by native-state H/D exchange.[20] The clusters of cooperatively unfolding residues are shown in orange and purple. In PUF3 (orange and purple), the amides of both clusters are protected against exchange and the remainder of the protein is unfolded. In PUF4, only the amides of the purple cluster are protected and all other parts of the protein are unfolded.

These ordered structures are populated about 10% of the time and are distinctively different from the random coil structures that are populated during the largest fraction of the time. The four regions with transient structure in unfolded apoflavodoxin actually reflect a population of PUF3, whereas Helix-U2 reflects a population of PUF4. As both PUFs are excursions from the off-pathway species, secondary structure formation in unfolded apoflavodoxin indeed promotes nonproductive folding toward an off-pathway folding intermediate.

Implications for Protein Folding. The presence of non-native secondary structure elements in unfolded proteins is probably a widespread phenomenon. However, subsequent formation of folding intermediates that contain these non-native structure elements is likely but rarely reported. For example, in the case of β-lactoglobulin, an on-pathway intermediate with some fluctuating non-native helical structure near the N-terminus, which is later converted to β-sheet, is formed.[46] In the case of the N-domain of phosphoglycerate kinase, a non-native helical species is formed, but it is unknown whether this species is on- or off-pathway.[47]

An off-pathway intermediate plays a major role during *A. vinelandii* apoflavodoxin folding and is also observed during the kinetic folding of three other proteins with an $\alpha-\beta$ parallel topology of which the folding mechanism has been studied:[22] *Anabaena* apoflavodoxin,[48] *Fusarium solani pisi* cutinase,[49] and *E. coli* CheY.[50] In this study, it is proven for the first time that formation of native and non-native helices within an unfolded $\alpha-\beta$ parallel protein and subsequent helix docking leads to formation of a compact helical off-pathway intermediate. Three relatively stable helices are populated in unfolded apoflavodoxin and dock non-natively onto one another (Figure 6c) through transient interactions between their hydrophobic sides (Figure 6d). The non-native docking of these helices prevents formation of the parallel β-sheet of native protein. To produce native $\alpha-\beta$ parallel protein molecules, the off-pathway species needs to unfold and as a result non-native interactions and non-native secondary structure are disrupted.

The presented study shows that acquisition of nativelike topology is not necessarily the general result of the initial collapse in protein folding. Rather than directing productive folding, conformational preorganization in the unfolded state of an $\alpha-\beta$ parallel type protein promotes off-pathway species formation. During kinetic folding of this type of protein, helices are formed much more easily and thus more rapidly than sheets, especially as a parallel β-sheet is involved. Formation of helices can occur on the nanosecond time scale.[51] This is due to the highly local character of interactions in helices, whereas the residues that have to be brought into contact to form a parallel β-sheet are separated by many residues from one another.[51] Our observations indicate that especially proteins that contain domains with an $\alpha-\beta$ parallel topology seem susceptible to off-pathway intermediate formation.

Acknowledgment. The Netherlands Organization for Scientific Research supported this work. NMR spectra were recorded at the Utrecht Facility for High-Resolution NMR, The Netherlands. We thank Simon Lindhoud for generating the Gln48Cys/Cys69Ala protein and Bregje de Kort for preparing NMR samples.

Supporting Information Available: Relaxation data of apoflavodoxin unfolded in 3.4 M GuHCl plotted as a function of residue number, and additional support for the presence of hydrophobic tertiary interactions between residual helices in unfolded apoflavodoxin. Acquisition parameters used for NMR experiments with apoflavodoxin unfolded in 6.0 or 3.4 M GuHCl, and backbone resonance assignment of *A. vinelandii* apoflavodoxin in 6.0 or 3.4 M GuHCl. This material is available free of charge via the Internet at http://pubs.acs.org.

JA803841N

(46) Kuwata, K.; Shastry, R.; Cheng, H.; Hoshino, M.; Batt, C. A.; Goto, Y.; Roder, H. *Nat. Struct. Biol.* **2001**, *8*, 151–155.
(47) Reed, M. A.; Jelinska, C.; Syson, K.; Cliff, M. J.; Splevins, A.; Alizadeh, T.; Hounslow, A. M.; Staniforth, R. A.; Clarke, A. R.; Craven, C. J.; Waltho, J. P. *J. Mol. Biol.* **2006**, *357*, 365–372.
(48) Fernandez-Recio, J.; Genzor, C. G.; Sancho, J. *Biochemistry* **2001**, *40*, 15234–15245.
(49) Melo, E. P.; Chen, L.; Cabral, J. M.; Fojan, P.; Petersen, S. B.; Otzen, D. E. *Biochemistry* **2003**, *42*, 7611–7617.
(50) Lopez-Hernandez, E.; Cronet, P.; Serrano, L.; Munoz, V. *J. Mol. Biol.* **1997**, *266*, 610–620.
(51) Bieri, O.; Kiefhaber, T. *Biol. Chem.* **1999**, *380*, 923–929.