J|A|C|S ARTICLES

Published on Web 01/26/2009

Noncooperative Formation of the Off-Pathway Molten Globule during Folding of the α−β Parallel Protein Apoflavodoxin

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 November 24, 2008; E-mail: carlo.vanmierlo@wur.nl

Abstract: During folding of many proteins, molten globules are formed. These partially folded forms of proteins have a substantial amount of secondary structure but lack virtually all tertiary side-chain packing characteristic of native structures. Molten globules are ensembles of interconverting conformers and are prone to aggregation, which can have detrimental effects on organisms. Consequently, molten globules attract considerable attention. The molten globule that is observed during folding of flavodoxin from *Azotobacter vinelandii* is a kinetically off-pathway species, as it has to unfold before the native state of the protein can be formed. This intermediate contains helices and can be populated at equilibrium using guanidinium hydrochloride as denaturant, allowing the use of NMR spectroscopy to follow molten globule formation at the residue level. Here, we track changes in chemical shifts of backbone amides, as well as disappearance of resonances of unfolded apoflavodoxin, upon decreasing denaturant concentration. Analysis of the data shows that structure formation within virtually all parts of the unfolded protein precedes folding to the molten globule state. This folding transition is noncooperative and involves a series of distinct transitions. Four structured elements in unfolded apoflavodoxin transiently interact and subsequently form the ordered core of the molten globule. Although hydrophobic, tryptophan side chains are not involved in the latter process. This ordered core is gradually extended upon decreasing denaturant concentration, but part of apoflavodoxin's molten globule remains random coil in the denaturant range investigated. The results presented here, together with those reported on the molten globule of α-lactalbumin, show that helical molten globules apparently fold in a noncooperative manner.

Introduction

Folding of proteins to conformations with proper biological activities is of vital importance for all living organisms. To describe folding, the concept of a multidimensional energy landscape or folding funnel arose from a combination of experiment, theory and simulation.[1−5] In this model, unfolded protein molecules descend along a funnel describing the free energy of folding, until the folding molecules 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.

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. When the folding species is on-pathway, it is productive for folding. In contrast, when the species is off-pathway, it is trapped as such that the native structure cannot be reached without substantial reorganizational events.[6] A decrease of the folding rate due to the presence of off-pathway intermediates, which are kinetically trapped and partially folded, increases the likelihood of protein aggregation. This aggregation can have detrimental effects on organisms. Folding intermediates are involved during folding of most proteins studied to date.[6] Kinetic intermediates that appear early during folding have been shown to resemble the relatively stable molten globule intermediates, found for several proteins under mildly denaturing conditions.[7] This resemblance suggests that these molten globules can be considered as models of transient intermediates.[6]

The term "molten globule" was introduced to describe compact partially folded ensembles of species that have a substantial amount of secondary structure, but lack virtually all tertiary side-chain packing characteristic of native structures.[7−10] The close similarity between molten globule states at equilibrium and those formed during the early stages of refolding has been demonstrated for α-lactalbumin,[11−13] apomyoglobin,[14,15] RNase H,[16] T4 lysozyme[17] and Im7.[18] Although often poorly defined,

(1) Bryngelson, J. D.; Onuchic, J. N.; Socci, N. D.; Wolynes, P. G. *Proteins* **1995**, *21*, 167–195.
(2) Dill, K. A.; Chan, H. S. *Nat. Struct. Biol.* **1997**, *4*, 10–19.
(3) Dinner, A. R.; Sali, A.; Smith, L. J.; Dobson, C. M.; Karplus, M. *Trends Biochem. Sci.* **2000**, *25*, 331–339.
(4) Vendruscolo, M.; Paci, E.; Dobson, C. M.; Karplus, M. *Nature* **2001**, *409*, 641–645.
(5) Fersht, A. R.; Daggett, V. *Cell* **2002**, *108*, 573–582.
(6) Jahn, T. R.; Radford, S. E. *Arch. Biochem. Biophys.* **2008**, *469*, 100–117.
(7) Arai, M.; Kuwajima, K. *Adv. Protein Chem.* **2000**, *53*, 209–282.
(8) Ohgushi, M.; Wada, A. *FEBS Lett.* **1983**, *164*, 21–24.
(9) Ptitsyn, O. B. *Adv. Protein Chem.* **1995**, *47*, 83–229.
(10) Redfield, C. *Methods* **2004**, *34*, 121–132.
(11) Arai, M.; Kuwajima, K. *Fold Des.* **1996**, *1*, 275–287.
(12) Forge, V.; Wijesinha, R. T.; Balbach, J.; Brew, K.; Robinson, C. V.; Redfield, C.; Dobson, C. M. *J. Mol. Biol.* **1999**, *288*, 673–688.
(13) Arai, M.; Ito, K.; Inobe, T.; Nakao, M.; Maki, K.; Kamagata, K.; Kihara, H.; Amemiya, Y.; Kuwajima, K. *J. Mol. Biol.* **2002**, *321*, 121–132.

molten globule-like intermediates have exposed hydrophobic groups and are prone to forming protein aggregates via specific interactions between structured elements.[19] Inside a living cell ample opportunities exist for aggregation of partially folded protein molecules and this nonproductive aggregation can compete with productive folding.[20,21] Detailed characterization of various equilibrium molten globule states promises to provide valuable insights into the transient kinetic molten globules that are important in folding pathways and in protein aggregation processes.

Here, NMR spectroscopy is used to report the folding at the residue-level of the molten globule-like folding intermediate of flavodoxin, a 179-residue protein from *Azotobacter vinelandii*. Flavodoxins are monomeric proteins involved in electron transport, and contain a noncovalently bound flavin mononucleotide (FMN) cofactor.[22] The proteins consist of a single structural domain and adopt the flavodoxin-like or α-β parallel topology, which is widely prevalent in nature. Both denaturant-induced equilibrium and kinetic (un)folding of flavodoxin and apoflavodoxin (i.e., flavodoxin without FMN) have been characterized using guanidine hydrochloride (GuHCl) as denaturant.[23−28] The folding data show that apoflavodoxin autonomously folds to its native state, which is structurally identical to flavodoxin except that residues in the flavin-binding region of the apo protein have considerable dynamics.[29,30] In presence of FMN, binding of FMN to native apoflavodoxin is the last step in flavodoxin folding.

The kinetic folding of apoflavodoxin involves an energy landscape with 2 folding intermediates and is described by: I_{off} $\leftrightarrow$ unfolded apoflavodoxin $\leftrightarrow$ I_{on} $\leftrightarrow$ native apoflavodoxin.[25] Intermediate I_{on} is an obligatory species on the productive route from unfolded to native protein. I_{on} is highly unstable and is therefore 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 thus acts as a trap.[25] The equilibrium unfolding of apoflavodoxin is described by: I_{off} $\leftrightarrow$ unfolded apoflavodoxin $\leftrightarrow$ native apoflavodoxin (Figure 1).[25] Intermediate I_{off} populates significantly in the concentration range of 1 to 3 M GuHCl (Figure 1d). The off-pathway species is molten globule-like: it is compact, it lacks the characteristic structure of native apoflavodoxin, its three tryptophans are solvent exposed in contrast to the situation in native protein, it contains helices and has severely broadened NMR resonances due to exchange between different conformers on the micro- to millisecond time scale.[25,31,32] Elevated protein concentrations[31] and macromolecular crowding[32] cause severe aggregation of this species. The formation of an off-pathway species is typical of proteins with a flavodoxin-like topology.[27]

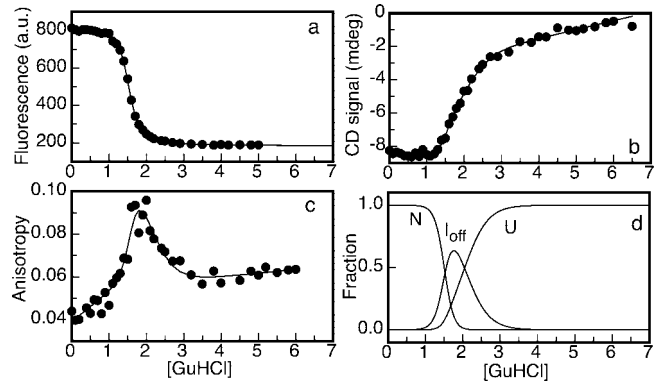

Figure 1. Guanidine hydrochloride-induced equilibrium (un)folding of *A. vinelandii* apoflavodoxin.[25] (a) Fluorescence emission at 340 nm with excitation at 280 nm. (b) Circular dichroism at 222 nm. The transition midpoints of the denaturation curves obtained by CD and fluorescence spectroscopy do not coincide, which is characteristic for the population of an intermediate. (c) Fluorescence anisotropy data detected with a 335 nm cutoff filter, excitation is at 300 nm. The curve is biphasic due to population of an intermediate. The solid lines in panels a to c are the result of a global fit of a three-state model for equilibrium (un)folding (I_{off} $\leftrightarrow$ Unfolded apoflavodoxin $\leftrightarrow$ Native apoflavodoxin).[25] (d) Normalized population of native (N), off-pathway intermediate (I_{off}), and unfolded (U) apoflavodoxin molecules as a function of denaturant concentration. The protein concentration is 5.6 μM in 100 mM potassium pyrophosphate, pH 6.0, and the data are recorded at 25 °C.

Backbone amide resonances of proteins can be followed in a series of $^{1}H-^{15}N$ heteronuclear single quantum coherence (HSQC) spectra acquired at different denaturant concentrations, as was originally done to study the folding of the molten globule state of α-lactalbumin.[33,34] Previous experiments, in which this HSQC-based method was used, showed that native apoflavodoxin unfolds highly cooperatively upon addition of denaturant[31] and also indicated that the off-pathway intermediate potentially folds in a noncooperative manner. However, molecular details of the formation of I_{off} could not be revealed by NMR spectroscopy, because the assignments of NMR resonances of the backbone amides of unfolded apoflavodoxin were lacking.

Recently, heteronuclear NMR spectroscopy has been used to assign resonances of unfolded apoflavodoxin and to characterize the conformational and dynamic properties of the unfolded protein at the residue level.[35] The study shows that the unfolded protein at 6.0 M GuHCl behaves as a random coil. However, at 3.4 M denaturant, unfolded apoflavodoxin has four transiently

(14) Hughson, F. M.; Wright, P. E.; Baldwin, R. L. *Science* **1990**, *249*, 1544–1548.
(15) Jennings, P. A.; Wright, P. E. *Science* **1993**, *262*, 892–6.
(16) Raschke, T. M.; Marqusee, S. *Nat. Struct. Biol.* **1997**, *4*, 298–304.
(17) Kato, H.; Feng, H.; Bai, Y. *J. Mol. Biol.* **2007**, *365*, 870–880.
(18) Spence, G. R.; Capaldi, A. P.; Radford, S. E. *J. Mol. Biol.* **2004**, *341*, 215–226.
(19) Jaenicke, R.; Seckler, R. *Adv. Protein Chem.* **1997**, *50*, 1–59.
(20) Ellis, R. J.; Hartl, F. U. *Curr. Opin. Struct. Biol.* **1999**, *9*, 102–110.
(21) Hartl, F. U.; Hayer-Hartl, M. *Science* **2002**, *295*, 1852–1858.
(22) Mayhew, S. G.; Tollin, G. General properties of flavodoxins. In *Chemistry and biochemistry of flavoenzymes*; Muller, F., Ed.; CRC Press: Boca Raton, FL, 1992; Vol. 3, pp 389−426.
(23) van Mierlo, C. P.; van Dongen, W. M.; Vergeldt, F.; van Berkel, W. J.; Steensma, E. *Protein Sci.* **1998**, *7*, 2331–2344.
(24) van Mierlo, C. P.; Steensma, E. *J. Biotechnol.* **2000**, *79*, 281–298.
(25) Bollen, Y. J.; Sanchez, I. E.; van Mierlo, C. P. *Biochemistry* **2004**, *43*, 10475–10489.
(26) Bollen, Y. J.; Nabuurs, S. M.; van Berkel, W. J.; van Mierlo, C. P. *J. Biol. Chem.* **2005**, *280*, 7836–7844.
(27) Bollen, Y. J.; van Mierlo, C. P. *Biophys Chem* **2005**, *114*, 181–189.
(28) Bollen, Y. J.; Kamphuis, M. B.; van Mierlo, C. P. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 4095–4100.
(29) 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.
(30) Steensma, E.; van Mierlo, C. P. *J. Mol. Biol.* **1998**, *282*, 653–666.
(31) van Mierlo, C. P.; van den Oever, J. M.; Steensma, E. *Protein Sci.* **2000**, *9*, 145–157.
(32) Engel, R.; Westphal, A. H.; Huberts, D. H.; Nabuurs, S. M.; Lindhoud, S.; Visser, A. J.; van Mierlo, C. P. *J. Biol. Chem.* **2008**, *283*, 27383–27394.
(33) Schulman, B. A.; Kim, P. S.; Dobson, C. M.; Redfield, C. *Nat. Struct. Biol.* **1997**, *4*, 630–634.
(34) Redfield, C.; Schulman, B. A.; Milhollen, M. A.; Kim, P. S.; Dobson, C. M. *Nat. Struct. Biol.* **1999**, *6*, 948–952.
(35) Nabuurs, S. M.; Westphal, A. H.; van Mierlo, C. P. *J. Am. Chem. Soc.* **2008**, *130*, 16914–16920.

ordered regions. These regions have restricted flexibility on the (sub)nanosecond time scale. Three of these regions transiently form α-helices and the fourth ordered region transiently adopts non-native structure, which is neither α-helix nor β-strand.[35] These ordered segments transiently dock non-natively, causing nonproductive folding toward the off-pathway folding intermediate I_{off}.

Because NMR assignments of GuHCl-unfolded apoflavodoxin recently became available,[35] the above-mentioned HSQC-based method can now be used to follow formation of I_{off} at the residue level, as is reported here. At denaturant concentrations that range from 4.05 to 1.58 M GuHCl, 18 $^{1}H-^{15}N$ HSQC spectra of unfolded apoflavodoxin were acquired. Subsequently, the chemical shifts of the amides of the unfolded protein were determined in these spectra. In addition, the denaturant-dependent disappearance of the amide cross peaks of unfolded apoflavodoxin were tracked. Analysis of the results obtained shows that formation of the molten globule off-pathway species occurs noncooperatively and is preceded by formation of structured regions in unfolded apoflavodoxin.

Experimental Section

Protein Expression, Purification, and 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. This protein variant is largely similar to wild-type flavodoxin regarding both redox potential of holoprotein and stability of apoprotein.[23,36] Uniformly $^{13}C-^{15}N$ labeled flavodoxin was obtained from transformed *Escherichia coli* cells grown on $^{13}C-^{15}N$ labeled algae medium (Silantes, Germany), and purified as described.[23]

Unfolded apoflavodoxin was obtained by denaturing flavodoxin in 6 M guanidine hydrochloride (GuHCl, ultrapure from Sigma). Subsequently, FMN was removed via gel filtration at 7 M GuHCl. Two stock solutions of 100 μM protein were prepared in 10% D_2O with 2,2-dimethyl-2-silapentane-5-sulfonic acid (DSS) as internal chemical shift reference. GuHCl concentrations of the stock solutions were 0.06 and 4.05 M GuHCl, respectively. Subsequently, 550 μL NMR samples of apoflavodoxin at various denaturant concentrations were prepared by appropriately mixing both stock solutions. In total, 18 apoflavodoxin samples with GuHCl concentrations ranging from 4.05 to 1.58 M GuHCl were prepared. After NMR experiments were finished, refractometry was used to determine the GuHCl concentration in each individual sample.[37] The buffer used in all experiments was 100 mM potassium pyrophosphate (Sigma), pH 6.0.

NMR Spectroscopy. Spectra were recorded on a Bruker Avance 700 MHz machine equipped with a triple-resonance 5 mm inverse probe with a self-shielded z-gradient. Sample temperature was 25 °C. Gradient-enhanced $^{1}H-^{15}N$ HSQC spectra[38,39] were recorded of apoflavodoxin samples at different GuHCl concentrations. Spectra were acquired at 4.05, 3.61, 3.41, 3.31, 3.08, 2.94, 2.76, 2.65, 2.55, 2.49, 2.39, 2.25, 2.16, 2.02, 1.92, 1.79, 1.68, or 1.58 M GuHCl. In the ^{1}H dimension of the HSQC experiments, 1024 complex data points were acquired, whereas in the indirect ^{15}N dimension 512 complex data points were collected. Spectral widths were 6010 and 1750 Hz in t_2 and t_1, respectively. With the number of scans set to 16, each HSQC experiment lasted 2 h and 45 min.

To assign the proton resonances of the indole $N^{\epsilon}H$ groups of the three tryptophans of unfolded apoflavodoxin, use was made of three tryptophan variants of apoflavodoxin. These variants contained a single tryptophan residue (W74), or either the W74/W128 or W74/W167 pair, which were generated through replacement of tryptophans by phenylalanines using oligo-directed mutagenesis.[40] One-dimensional proton NMR spectra were recorded of all three tryptophan variants of apoflavodoxin unfolded in 5 M GuHCl.

Determination of Cross Peak Volumes. All spectra were processed with NMRPipe[41] and analyzed with NMRviewJ.[42] Data were zero-filled to 2048 points in both dimensions to gain resolution and smooth the shape of the peaks. The volume of a HSQC cross peak was calculated by summing the intensities of all points within an elliptical region that has its center at the maximum of the cross peak. The line widths at the cross peak's half-height in the ^{1}H and the ^{15}N dimension, respectively, were multiplied by a factor of 1.37 to obtain the values of the radii that define the elliptical region used.[43]

Whereas adequate sample shimming is easily achieved, optimal tuning and matching of the probe of the 700 MHz instrument was not possible due to the effects of GuHCl present in the samples. Upon increasing denaturant concentration, the accompanying reduction in NMR receiver coil quality factor Q leads to additional loss of intensity of all signals.[31] A proper correction for these losses is achieved by using HSQC cross peak volumes of backbone amides of residues that behave random coil-like in the range of 4.05 to 1.58 M GuHCl. The denaturant-dependencies of these cross peak volumes showed no folding transitions due to formation of intermediate I_{off}. Instead, the corresponding chemical shifts change nearly linearly in this denaturant concentration range. Nine residues of unfolded apoflavodoxin showing this behavior were selected (i.e., Arg15, Lys16, Ala18, Lys22, Lys23, Arg24, Glu28, Thr29, and Ser31). The cross peak volumes of the corresponding backbone amides are directly related to the number of nonnative apoflavodoxin molecules (i.e., unfolded and I_{off}) present in the NMR sample. At each denaturant concentration j these nine cross peak volumes were averaged, and the resulting averaged cross peak volume $\langle V\rangle_j$ was used to properly correct the data for the discussed NMR signal losses. The backbone amide cross peak volume of residue i at denaturant concentration j, i.e., $V_{i,j}$, was divided by $\langle V\rangle_j$ to obtain the corrected cross peak volume $V_{i,j}^{corr}$ according to:

$$V_{i,j}^{corr} = \frac{V_{i,j}(1-f_{N,j})}{\langle V\rangle_j} \quad (1)$$

in which $f_{N,j}$ is the fraction of apoflavodoxin molecules that is native at GuHCl concentration j. The $(1 - f_{N,j})$ term is required to correct for the presence of native apoflavodoxin at GuHCl concentrations below 2 M. These native protein molecules do not contribute to the NMR signals of the unfolded state and cause a further decrease of these signals upon lowering denaturant concentration. Fraction $f_{N,j}$ (Figure 1) is calculated using GuHCl-induced equilibrium (un)folding data of apoflavodoxin.[25]

Determination of Midpoints of Unfolding at the Residue Level. The disappearance of each individual HSQC cross peak of unfolded apoflavodoxin upon decreasing denaturant concentration could be analyzed according to a two-state model of (un)folding. When a residue is highly flexible, as is the case in an unfolded protein, it gives rise to a very sharp cross peak in a specific region of the HSQC spectrum,[35] making this analysis possible. Upon folding, this particular cross peak disappears as backbone motions become restricted.

(36) Steensma, E.; Heering, H. A.; Hagen, W. R.; van Mierlo, C. P. M. *Eur. J. Biochem.* **1996**, *235*, 167–172.
(37) Nozaki, Y., The preparation of guanidine hydrochloride. In *Methods in enzymology*; Hirs, C. H. W., Timasheff, S. N., Eds.; Academic Press: New York, 1972; Vol. 26, pp 43−50.
(38) Palmer, A. G., 3rd; Cavanagh, J.; Wright, P. E.; Rance, M. *J. Magn. Reson.* **1991**, *93*, 151–170.
(39) Kay, L. E.; Keifer, P.; Saarinen, T. *J. Am. Chem. Soc.* **1992**, *114*, 10663–10665.
(40) Visser, N. V.; Westphal, A. H.; van Hoek, A.; van Mierlo, C. P. M.; Visser, A. J. W. G.; van Amerongen, H. *Bioph. J.* **2008**, *95*, 2462–2469.
(41) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. *J. Biomol. NMR* **1995**, *6*, 277–293.
(42) Johnson, B. A.; Blevins, R. A. *J. Biomol.NMR* **1994**, *4*, 603–614.
(43) Rischel, C. *J. Magn. Res. Ser. A* **1995**, *116*, 255–258.

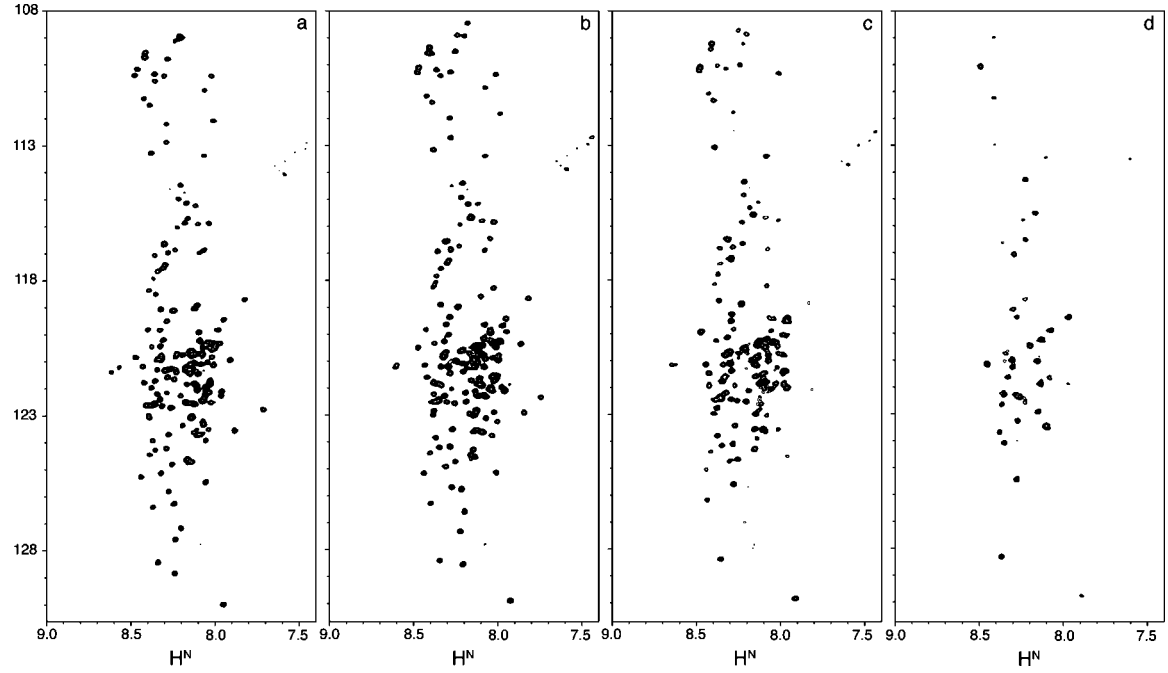

Figure 2. NMR resonances of unfolded apoflavodoxin disappear upon folding of the protein. Shown are ^{1}H−^{15}N HSQC spectra of apoflavodoxin in (a) 4.05 M GuHCl, (b) 3.08 M GuHCl, (c) 2.39 M GuHCl, and (d) 1.58 M GuHCl. Protein concentration is 100 μM in 100 mM potassium pyrophosphate, pH 6.0, 25 °C.

The free energy of (un)folding is assumed to be linearly dependent on denaturant concentration,[44] according to:

$$\Delta G_{U-F} = \Delta G^0_{U-F} + m[D] \quad (2)$$

with ΔG_{U-F} being the difference in free energy between the unfolded state and a folded state at a particular denaturant concentration. ΔG^0_{U-F} is ΔG_{U-F} at zero denaturant concentration, m the slope at the midpoint of the unfolding transition, and [D] the concentration of denaturant, respectively. It follows from eq 2 that at C_m:[45]

$$\Delta G^0_{U-F} = -m \cdot C_m \quad (3)$$

Thus:

$$\Delta G_{U-F} = m([D] - C_m) \quad (4)$$

Incorporation of eq 4 into the linear extrapolation method,[46] which takes into account the linear dependence of the pre- and postunfolding data on the denaturant concentration, leads to the following general equation:

$$Y_{obs} = \frac{(\alpha_F + \beta_F[D]) + (\alpha_U + \beta_U[D])\exp\{-[m([D] - C_m)/RT]\}}{1 + \exp\{-[m([D] - C_m)/RT]\}} \quad (5)$$

with Y_{obs} representing any physical parameter that characterizes the folded and unfolded states of a protein at a particular denaturant concentration, α and β are intercepts and slopes of pre- and postunfolding regions, and RT is 0.59 kcal mol^{-1}.

In case of the apoflavodoxin folding data, obtained by a series of denaturant-dependent HSQC spectra, cross peaks that solely correspond with highly flexible residues were followed. As there is no interference of these cross peaks with those arising from residues that have restricted motions due to folding, α_F and β_F can be set to 0 in eq 5, resulting in:

$$Y_{obs} = \frac{(\alpha_U + \beta_U[D])\exp\{-[m([D] - C_m)/RT]\}}{1 + \exp\{-[m([D] - C_m)/RT]\}} \quad (6)$$

The residue-specific C_m values were calculated by applying eq 6 to the denaturant-dependent corrected cross peak volumes $V^{corr}_{i,j}$ of 68 residues of unfolded apoflavodoxin. If a cross peak was no longer observed above noise level, the corresponding cross peak volume was set to zero. Cross peak volumes $V^{corr}_{i,j}$ at 0, 0.3, 0.6, and 0.9 M GuHCl of these 68 residues were set to zero, as no highly flexible residues typical for unfolded protein were experimentally detected at these denaturant concentrations.[31] Indeed, the GuHCl-induced equilibrium (un)folding data of apoflavodoxin show that the protein is native up to 0.9 M GuHCl.[25]

Fitting of Denaturant-Dependent Chemical Shift Changes. Chemical shifts of 114 backbone amides of unfolded apoflavodoxin could be followed in the 18 HSQC spectra obtained in the denaturant range of 4.05 to 1.58 M GuHCl. Depending on how ^{1}H and ^{15}N chemical shifts change upon altering denaturant concentration, either a linear, exponential or polynomial function was used to fit to these data. Subsequently, for each NH group at a particular denaturant concentration [D], the magnitude of the nonlinear dependence of chemical shifts on denaturant concentration ($X_{[D]}$) was calculated according to:

$$X_{[D]} = \left|\frac{d^2(\delta^1H)}{d^2[D]}\right| \cdot 10 + \left|\frac{d^2(\delta^{15}N)}{d^2[D]}\right| \quad (7)$$

in which δ^1H is the proton chemical shift and δ^{15}N the nitrogen chemical shift. When $X_{[D]}$ is above zero, the chemical shift of either ^{1}H or ^{15}N, or of both nuclei, has a nonlinear dependency on the denaturant concentration.

Results and Discussion

Residue-Specific Folding of Apoflavodoxin's Molten Globule Can Be Followed by NMR Spectroscopy. Gradient-enhanced ^{1}H−^{15}N HSQC spectra were recorded of 18 apofla-

(44) Pace, C. N. *Methods Enzymol.* **1986**, *131*, 266–280.
(45) Jackson, S. E.; Moracci, M.; elMasry, N.; Johnson, C. M.; Fersht, A. R. *Biochemistry* **1993**, *32*, 11259–11269.
(46) Santoro, M. M.; Bolen, D. W. *Biochemistry* **1988**, *27*, 8063–8068.

vodoxin samples with GuHCl concentrations ranging from 4.05 to 1.58 M GuHCl (Figure 2). In this denaturant range, the transition of unfolded protein to I_{off} should be detectable (Figure 1d). Due to the relatively low protein concentration used, that is, 100 μM, no protein aggregation is observed in these NMR samples. In the HSQC spectra, cross peaks of unfolded apoflavodoxin were assigned using the assignments of all H^N and ^{15}N resonances of unfolded apoflavodoxin in 6.0 M and in 3.4 M GuHCl[35] and by using a 3D HNCA spectrum of apoflavodoxin in 2.22 M GuHCl.

In the denaturant range from 4.0 to 1.6 M GuHCl, exchange between unfolded apoflavodoxin and I_{off} is slow on the NMR chemical shift time scale (i.e., $\sim$10 s^{-1} at 4.0 M GuHCl and $\sim$60 s^{-1} at 1.6 M GuHCl).[25] Consequently, upon decreasing the denaturant concentration, cross peaks arising from unfolded apoflavodoxin should disappear from the corresponding HSQC spectra, and cross peaks arising from I_{off} should appear. However, as I_{off} is a molten globule, most of its resonances are not observed because these resonances are broadened beyond detection.[31] This severe line broadening is due to conformational exchange processes on the micro- to millisecond time scale between the ensemble of conformers that represents the molten globule.[10]

At 1.58 M GuHCl, the lowest denaturant concentration used, 34% of the total number of protein molecules is native (Figure 1d).[25] At this denaturant concentration, conformational exchange between native and unfolded apoflavodoxin is slow on the NMR chemical shift time scale.[31] However, due to the low protein concentration used, at 1.58 M GuHCl no backbone amide cross peaks are observed for native apoflavodoxin in the corresponding HSQC spectrum (except for the highly flexible C-terminal Leu179). In contrast, cross peaks arising from unfolded apoflavodoxin, which has large intrinsic flexibility,[35] are much sharper and thus visible.

In conclusion, upon decreasing GuHCl concentration from 4.05 to 1.58 M, HSQC spectra show almost exclusively cross peaks arising from unfolded apoflavodoxin. Tracking the denaturant-dependent changes in the corresponding cross peak volumes and chemical shifts enables the study of the residue-specific folding of the unfolded protein to I_{off}.

Several Residues Have Comparable Dynamical and Conformational Properties in I_{off} and Unfolded Apoflavodoxin. Figure 2 shows HSQC spectra of apoflavodoxin in different GuHCl concentrations. The HSQC spectrum obtained at 4.05 M GuHCl (Figure 2a) is typical for an unfolded protein and has limited dispersion in the ^1H dimension, whereas dispersion in the ^{15}N dimension is good. All cross peaks observed at 4.05, 3.61, 3.41, 3.31, 3.08, and 2.94 M GuHCl are sharp, as expected for an unfolded protein. Upon decreasing the GuHCl concentration from 4.05 to 2.94 M, the number of cross peaks detected remains constant and several cross peaks shift. Upon decreasing the denaturant concentration to 2.76 M GuHCl, the volume of some cross peaks reduces. The latter phenomenon becomes more pronounced upon further decreasing denaturant concentration, and at 2.48 M GuHCl a few cross peaks have disappeared from the HSQC spectrum. At 1.58 M GuHCl, many more cross peaks are not observed in the corresponding HSQC spectrum (Figure 2d).

The majority of backbone amide cross peaks of unfolded apoflavodoxin disappears upon decreasing the denaturant concentration from 4.05 to 1.58 M GuHCl. However, 27 sharp cross peaks with considerable intensity are still present at 1.58 M GuHCl (Figure 2d) and arise from the following residues: Phe6, Lys13−Lys16, Ala18, Lys19, Ile21−Asp32, Leu34, Val36, Gly60, Asp68−Glu70, and Ser178. These 27 cross peaks exhibit only small chemical shift changes upon decreasing the denaturant concentration. Equilibrium unfolding studies show that at 1.58 M GuHCl, 34% of the protein molecules is native, 53% is off-pathway folding intermediate and only 13% is unfolded (Figure 1).[25] Consequently, at 1.58 M GuHCl, the 27 residues mentioned must have comparable dynamical and conformational properties in I_{off} and unfolded apoflavodoxin.

Unfolded Apoflavodoxin Folds Noncooperatively to I_{off}. To study the residue-specific folding of unfolded apoflavodoxin to I_{off}, denaturant-dependent changes in the $^1H-^{15}N$ cross peak volumes of unfolded protein are followed. Cross peak volumes instead of cross peak intensities are used to correct for changes in R_2 relaxation times, as several regions of unfolded apoflavodoxin become transiently ordered,[35] and as upon reducing GuHCl concentration the viscosity of the solution diminishes. Due to the presence of GuHCl in the apoflavodoxin samples, intensities of all NMR signals differ slightly between the different HSQC spectra.[31] A proper correction for these intensity changes, which are not due to protein folding, is possible and is applied to all cross peak volumes, as described in the Experimental Section.

Changes in corrected cross peak volumes of 68 well-resolved backbone amides and of the 3 tryptophan indole $N^{\epsilon}H$ groups of unfolded apoflavodoxin could be followed upon decreasing the denaturant concentration. Some examples of folding transitions observed are shown in Figure 3. A two-state model suffices to describe each folding transition.

The apparent midpoints of folding (i.e., C_m values, see the Experimental Section) of the backbone amides of 68 residues of unfolded apoflavodoxin are shown in Figure 4 and listed in Supporting Information, Table S1 (including C_m values of the tryptophan indole $N^{\epsilon}H$ groups). The residues investigated are distributed along the entire sequence of apoflavodoxin. Significant differences exist between C_m values, since they range from 1.5 to 2.7 M GuHCl, and five clusters of residues can be identified (Figure 4b), although the boundaries of these clusters are not sharply defined.

The slope associated with the folding transition from unfolded apoflavodoxin to I_{off} as observed by far-UV CD is rather shallow (Figure 1b). In contrast, each folding transition detected at the residue level by NMR spectroscopy has a much steeper slope (Figure 3). The residue-specific C_m values cover a broad range of denaturant concentrations. Folding toward I_{off} involves a series of distinct transitions and as a result far-UV CD reports a shallow folding transition. The HSQC data presented here show that folding of I_{off} occurs noncooperatively.

Upon Decreasing Denaturant Concentration, the Ordered Core of I_{off} Becomes Gradually Extended. At 3.4 M GuHCl, four transiently ordered regions are detected in unfolded apoflavodoxin (comprising residues 41−53, 72−83, 99−122, and 160−176). These regions transiently dock onto one another in a non-native manner, and this non-native docking promotes folding toward the helical off-pathway intermediate.[35]

The residue specific folding data presented here show that the residues that have the largest average C_m value (i.e., Group-1, $C_m^{average} = 2.64 \pm 0.05$ M GuHCl; red in Figure 4; typical examples Gly53 and Ala92, Figures 3c and d) roughly coincides with those residues that are transiently ordered in unfolded apoflavodoxin (gray bars in Figure 4a). Thus, upon decreasing denaturant concentration, assembly of these four structured

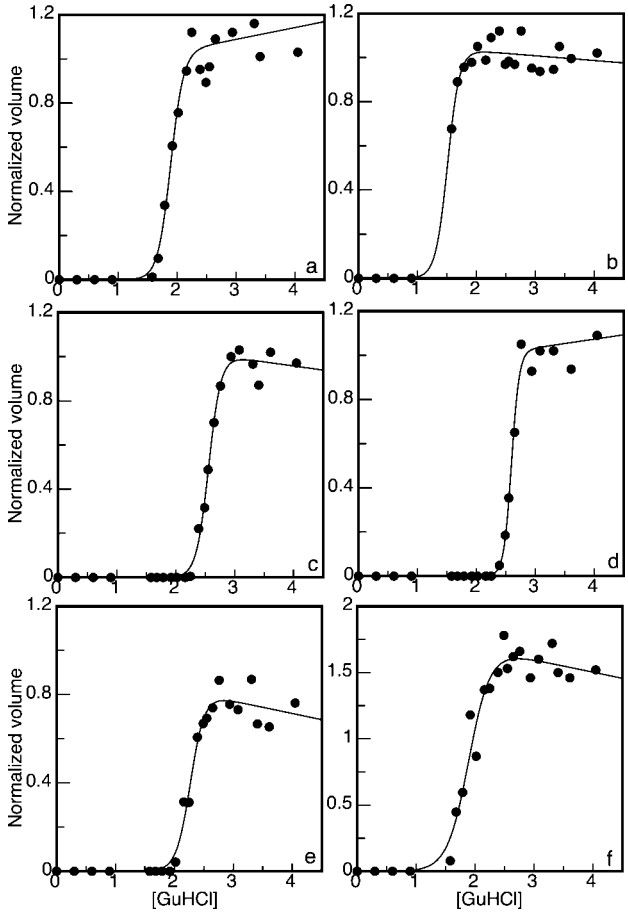

Figure 3. Folding of unfolded apoflavodoxin at the residue-level, as monitored by changes in cross peak volumes in $^{1}H-^{15}N$ HSQC spectra at various concentrations of GuHCl. Shown are cross peak volumes of the backbone amides of (a) Gly4, (b) Ala18, (c) Gly53, (d) Ala92, (e) Leu179, and (f) the indole $N^{\epsilon}H$ group of Trp128. All cross peaks are corrected for GuHCl-induced changes in volume that are not due to protein folding, as described in the Experimental Section. Cross peak volumes at 0, 0.3, 0.6, and 0.9 M GuHCl are set to zero, as no cross peaks typical for unfolded protein were detected at these denaturant concentrations. The solid curves show the fit of a two-state model of protein folding (i.e., eq 6, Experimental Section) to the apoflavodoxin folding data. The extracted midpoints of folding (i.e., C_m) are: 1.89 M (Gly4), 1.51 M (Ala18), 2.56 M (Gly53), 2.60 M (Ala92), 2.28 M (Leu179), and 1.90 M ($N^{\epsilon}H$ Trp128).

elements in unfolded apoflavodoxin leads to formation of the ordered core of I_{off}.

The C_m data of the group of residues that has the second largest average C_m value (i.e., Group-2, $C_m^{average} = 2.40 \pm 0.07$ M GuHCl) are highlighted in yellow in Figure 4. All corresponding residues are close in sequence to the residues of Group-1, which together form the core of I_{off}. This observation suggests that upon decreasing denaturant concentration the initial ordered core of I_{off} becomes extended.

A further reduction in denaturant concentration leads to additional sequential extension of the folded regions of I_{off}. This conclusion is drawn on basis of the C_m data of the group of residues that has the third largest average C_m value (i.e., Group-3, $C_m^{average} = 2.10 \pm 0.07$ M GuHCl, cyan in Figure 4). This distribution shows that most of the corresponding residues are close in sequence to those residues that already became ordered in I_{off} at higher denaturant concentrations. However, residues 7 to 9 cooperatively fold to newly formed structure in I_{off} and their folding does not extend existing ordered parts of I_{off}.

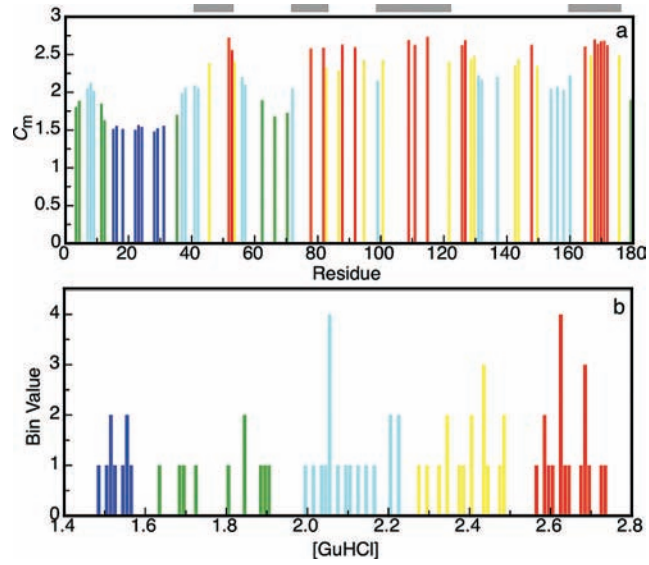

Figure 4. Midpoints of folding (C_m values) determined for the backbone amide groups and the three indole $N^{\epsilon}H$ groups of unfolded apoflavodoxin. (a) C_m values are plotted versus residue number, errors range between 0.002 and 0.075 M; each vertical bar is colored according to the coloring scheme of (b). Horizontal gray bars highlight the four transiently ordered regions in unfolded apoflavodoxin at 3.4 M GuHCl.[35] (b) Histogram of binned C_m values; the bin width is 0.01 M GuHCl. Five groups of residues are identified and are colored differently. Group-1 (C_m between 2.5 to 2.8 M GuHCl): red, Group-2 (C_m between 2.2 to 2.5 M GuHCl): yellow, Group-3 (C_m between 1.95 and 2.2 M GuHCl): cyan, Group-4 (C_m between 1.6 and 1.95 M GuHCl): green, and Group-5 (C_m between 1.4 and 1.6 M GuHCl): dark blue.

The fourth group of residues of unfolded apoflavodoxin has more widely distributed C_m values (i.e., Group-4, $C_m^{average} = 1.79 \pm 0.10$ M GuHCl, green in Figure 4; typical example Leu179, Figure 3e). Most of these residues reside in the N-terminal part of the protein and are positioned at the N- and C-terminal sides of ordered residues 7 to 9, thereby extending the structure formed by the latter residues.

Finally, NMR spectroscopy shows that the remaining fifth group of residues of unfolded apoflavodoxin has an average C_m value of 1.53 ± 0.03 M GuHCl (i.e., Group-5, dark blue in Figure 4; typical example Ala18, Figure 3b). This $C_m^{average}$ is within error identical to the C_m values derived from denaturant-induced unfolding of native apoflavodoxin detected by fluorescence emission spectroscopy ($C_m = 1.53 \pm 0.01$ M GuHCl, Figure 1a[25]) and by NMR spectroscopy ($C_m = 1.48 \pm 0.04$ M GuHCl[31]). Thus, the C_m values obtained for Arg15, Lys16, Ala18, Lys22, Lys23, Arg24, Glu28, Thr29, and Ser31 report folding of unfolded apoflavodoxin to native protein. Consequently, in the denaturant range investigated, these residues are unfolded in I_{off}.

Tryptophans Do Not Cause Folding of Unfolded Apoflavodoxin to I_{off}. Apoflavodoxin contains three tryptophans (i.e., Trp74, Trp128, and Trp167). Of these tryptophans, both Trp74 and Trp167 belong to regions with restricted flexibility in unfolded apoflavodoxin in 3.4 M GuHCl, as revealed by elevated ^{15}N R_2 relaxation rates of the corresponding backbone amides.[35] Analysis of the denaturant-dependent disappearance of the HSQC cross peaks arising from the indole $N^{\epsilon}H$ groups of these tryptophans of unfolded apoflavodoxin gives the following C_m values: Trp74-$N^{\epsilon}H$: 1.84 M, Trp128-$N^{\epsilon}H$: 2.28 M, and Trp167-$N^{\epsilon}H$: 2.38 M (Supporting Information Table S1). These values are significantly smaller than the average C_m

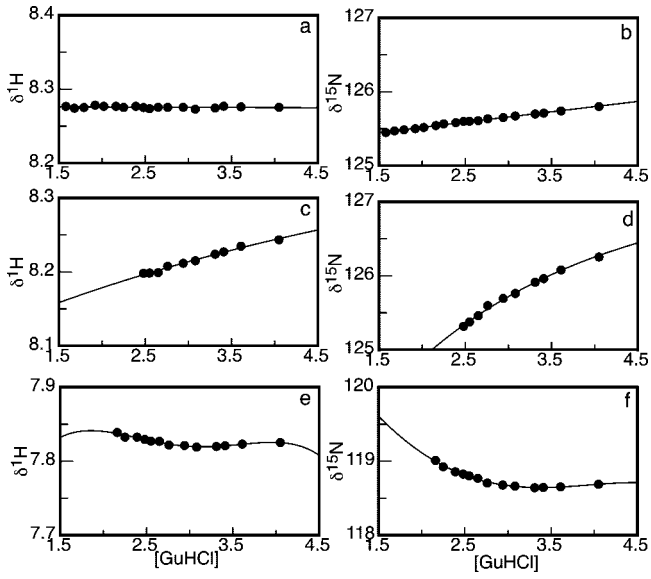

Figure 5. Denaturant-dependence of chemical shifts of various backbone amides of unfolded apoflavodoxin. Chemical shifts are shown as closed symbols (●) for (a) ^{1}H and (b) ^{15}N of the backbone amide of Lys22, (c) ^{1}H and (d) ^{15}N of the backbone amide of Leu52, and (e) ^{1}H and (f) ^{15}N of the backbone amide of Trp167. Lines show the fit of a linear (a, b), exponential (c, d) or polynomial function (e, f) to the data.

value of 2.64 ± 0.05 M that characterizes the folding of the residues of Group-1 (Figure 4).

As discussed, docking of the transiently ordered residues of Group-1 in unfolded apoflavodoxin leads to formation of the ordered core of I_{off}. The relatively low C_m values obtained for the $N^{\epsilon}H$ groups of the tryptophans of unfolded apoflavodoxin imply that the side chains of these amino acid residues are not involved in this process. Apoflavodoxin's tryptophans do not have substantial hydrophobic interactions with the ordered regions of the unfolded protein. In contrast, in case of lysozyme the hydrophobic tryptophan at position 62 plays an important role in folding, as it is involved in extensive long-range tertiary interactions in the urea-denatured state of the protein.[47] Apparently, tryptophans do not play a significant role in folding of I_{off}.

Transient Formation of Ordered Structure within Virtually All Parts of Unfolded Apoflavodoxin Precedes Transition to I_{off}. To further clarify the folding events that occur upon decreasing the GuHCl concentration, ^{1}H and ^{15}N chemical shifts of unfolded apoflavodoxin are followed. Changes in chemical shifts are good indicators for formation of ordered structure in unfolded proteins.[35,48] Chemical shifts of cross peaks of 114 backbone amides of unfolded apoflavodoxin have been followed in the 18 HSQC spectra discussed. These cross peaks were selected because they do not severely overlap with each other.

Chemical shifts of backbone amides that depend linearly on denaturant concentration (Figure 5a,b) merely reflect the change in average properties of the solvent surrounding these amides. Thus, no structure formation involving the corresponding residues takes place in unfolded apoflavodoxin. In contrast, chemical shifts of various backbone amides have a nonlinear dependence on denaturant concentration (Figure 5c−f), reflecting transient structure formation in unfolded apoflavodoxin molecules.

Depending on how the ^{1}H and ^{15}N chemical shifts change as a function of denaturant concentration, either a linear, exponential or polynomial function was used to fit the data. Subsequently, at a particular denaturant concentration [D] the magnitude of the nonlinear dependence of the chemical shift on denaturant concentration (i.e., $X_{[D]}$) is calculated per residue by taking the second derivatives of the corresponding fits (see the Experimental Section). $X_{[D]}$ is used as a measure of the increase in structure formation at the residue level in unfolded apoflavodoxin upon decreasing denaturant concentration. The $X_{[D]}$ data are presented in Figure 6.

Figure 6 shows that residues 21 to 35 of unfolded apoflavodoxin are characterized by $X_{[D]}$-values of almost zero at all denaturant concentrations used. Thus, no structure is formed in this region of unfolded apoflavodoxin. Note that these residues show no transition from unfolded apoflavodoxin to I_{off} in the denaturant range investigated. However, for all other residues, $X_{[D]}$ is well above zero at denaturant concentrations close to their C_m values. Consequently, these residues form transient structure in unfolded protein. Figure 6 shows that the first regions that become ordered upon decreasing denaturant concentration are the four regions that are transiently ordered in unfolded apoflavodoxin.[35] Non-native docking of these regions leads to formation of I_{off}. In addition, Figure 6 shows that upon decreasing GuHCl concentration, most of the remaining parts of the unfolded protein also become transiently ordered.

Clearly, virtually all residues of unfolded apoflavodoxin form transient structure upon decreasing denaturant concentration, and this process precedes the transition of the unfolded species to the off-pathway folding intermediate I_{off}.

Part of the off-Pathway Folding Intermediate I_{off} is Random Coil. As discussed, at 1.58 M GuHCl, the lowest denaturant concentration used in this study, 27 residues of apoflavodoxin have the dynamical and conformational properties observed for an unfolded protein. Most of these residues are clustered in two regions of apoflavodoxin; Lys13 to Val36 (Region-U1), and Gly60 to Glu70 (Region-U2). Determination of the denaturant-dependent changes in chemical shifts of the corresponding backbone amides (Figure 6), as well as the disappearance of the corresponding resonances (Figure 4a), enables the further characterization of these two regions.

Analysis of C_m values shows that at 1.58 M GuHCl Region-U1 must be unfolded in the ensemble of species that forms I_{off}, as no folding transition from unfolded protein to I_{off} is observed (Figure 4a; Group-5, dark blue). Figure 6 shows that in the range from 4.05 to 1.58 M GuHCl, part of Region-U1 (i.e., residues 21 to 36) has $X_{[D]}$ values close to or equal to zero. Thus, in this denaturant range no structure formation in the C-terminal part of Region-U1 takes place. Consequently, at 1.58 M GuHCl, the C-terminal part of Region-U1 behaves as a random coil. Whether this random coil behavior persists upon further decreasing denaturant concentration is unknown. The N-terminal part of Region-U1 has relatively small $X_{[D]}$ values in the denaturant range investigated. Consequently, at 1.58 M GuHCl, the N-terminal part of Region-U1 is involved in some minor structure formation. Upon further decreasing denaturant concentration, this protein part likely becomes structured in I_{off}.

The backbone amides of Region-U2 of apoflavodoxin are characterized by an average C_m value of 1.77 M GuHCl. The corresponding residues thus report the folding transition that

(47) 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.
(48) Dyson, H. J.; Wright, P. E. *Methods Enzymol.* **2005**, *394*, 299–321.

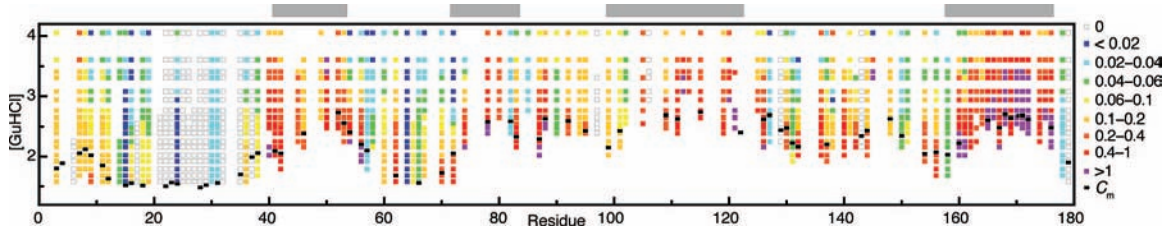

Figure 6. Magnitude $X_{[D]}$ of the nonlinear dependence of the ^1H and ^{15}N chemical shifts of unfolded apoflavodoxin on denaturant concentration versus residue number. The vertical axis shows the GuHCl concentration, and coloration indicates the magnitude of $X_{[D]}$ (see the Experimental Section). Horizontal gray bars highlight the four transiently ordered regions in unfolded apoflavodoxin.[35] The midpoints of unfolding determined for the backbone amide groups of unfolded apoflavodoxin (Figure 4a) are indicated by (-).

links unfolded apoflavodoxin with I_{off}. Therefore, these residues must be ordered in I_{off}. At 1.58 M GuHCl, which is still in the transition region of denaturant-induced equilibrium (un)folding of these residues, Region-U2 is unfolded in part of the ensemble of apoflavodoxin molecules. Consequently, at 1.58 M GuHCl, the corresponding backbone amide cross peaks are still observable in the HSQC spectrum. Upon further decreasing denaturant concentration, and in absence of denaturant, all residues of Region-U2 of apoflavodoxin will become ordered in I_{off}.

Molten Globules of Helical Nature Apparently Fold in a Noncooperative Manner. In unfolded apoflavodoxin several regions of the protein have reduced flexibility, which is due to transient helix formation and local and nonlocal hydrophobic interactions. Helices are the only regular secondary structure elements detected in unfolded apoflavodoxin and these helices are sufficiently stable to be present about 10% of the time.[35]

Upon protein folding, helices are formed much more rapidly than sheets, especially when parallel β-sheets are involved. This rapid helix formation is due to the highly local character of the interactions in helices, whereas the residues that have to be brought into contact to form a parallel β-sheet, as is the case for apoflavodoxin, are separated by many residues from one another.[49,50]

Rapid formation of three α-helices and transiently ordered structure that is neither α-helix nor β-strand and their subsequent non-native docking through hydrophobic interactions causes formation of the off-pathway intermediate during apoflavodoxin folding.[35] This docking of helices prevents formation of the parallel β-sheet of apoflavodoxin and causes the intermediate to be helical. The helical character of the intermediate is confirmed by far-UV CD measurements (to be published). Similar processes are expected to play a role during kinetic folding of other proteins with an α-β parallel topology,[35] causing them to be susceptible to off-pathway intermediate formation. Indeed, an off-pathway intermediate is experimentally observed for all other α-β parallel proteins of which the kinetic folding has been investigated: apoflavodoxin from *Anabaena*,[51] CheY,[52] cutinase,[53] and UMP/CMP kinase.[54]

Whereas previous NMR experiments showed that formation of native apoflavodoxin is highly cooperative,[31] the data presented here show that I_{off} is noncooperatively formed. This noncooperativity is probably due to the helical character of the off-pathway intermediate. Helices involve relatively short-range interactions and the folding of one helix may not affect the folding of other helices.[55] Indeed, in case of the α-helical domain of the molten globule of α-lactalbumin, which has native-like secondary structure, the α-helices fold at differing denaturant concentrations.[33] In contrast, the folding of the molten globule of human serum retinol-binding protein, which contains a significant amount of β-sheet structure, is significantly more cooperative. The formation of a stable β-sheet requires that at least two β-strands fold and interact, which is the cause of this cooperativity. Our observations concerning the folding of the off-pathway intermediate of apoflavodoxin, together with those of the molten globule of α-lactalbumin, show that the folding of helical molten globules is apparently noncooperative.

Acknowledgment. The Netherlands Organization for Scientific Research supported this work. NMR spectra were recorded at the Utrecht Facility for High-Resolution NMR, The Netherlands.

Supporting Information Available: Midpoints of unfolding (C_m values) determined for the backbone amide groups and the three $N^{\varepsilon}H$ indole groups of unfolded apoflavodoxin. This material is available free of charge via the Internet at http://pubs.acs.org.

JA8089476

(49) Bieri, O.; Kiefhaber, T. *Biol Chem* **1999**, *380*, 923–929.
(50) Plaxco, K. W.; Simons, K. T.; Baker, D. *J. Mol. Biol.* **1998**, *277*, 985–994.
(51) Fernandez-Recio, J.; Genzor, C. G.; Sancho, J. *Biochemistry* **2001**, *40*, 15234–15245.
(52) Kathuria, S. V.; Day, I. J.; Wallace, L. A.; Matthews, C. R. *J. Mol. Biol.* **2008**, *382*, 467–484.
(53) Otzen, D. E.; Giehm, L.; Baptista, R. P.; Kristensen, S. R.; Melo, E. P.; Pedersen, S. *Biochim. Biophys. Acta* **2007**, *1774*, 323–333.
(54) Lorenz, T.; Reinstein, J. *J. Mol. Biol.* **2008**, *381*, 443–455.
(55) Greene, L. H.; Wijesinha-Bettoni, R.; Redfield, C. *Biochemistry* **2006**, *45*, 9475–9484.