ELSEVIER

Biophysical Chemistry 114 (2005) 181–189

Biophysical Chemistry

http://www.elsevier.com/locate/biophyschem

Protein topology affects the appearance of intermediates during the folding of proteins with a flavodoxin-like fold

Yves J.M. Bollen[1], Carlo P.M. van Mierlo*

Department of Agrotechnology and Food Sciences, Laboratory of Biochemistry, Wageningen University, Dreijenlaan 3, NL-6703 HA Wageningen, The Netherlands

Received 3 June 2004; received in revised form 25 October 2004; accepted 8 December 2004
Available online 19 December 2004

Abstract

The topology of a native protein influences the rate with which it is formed, but does topology affect the appearance of folding intermediates and their specific role in kinetic folding as well? This question is addressed by comparing the folding data recently obtained on apoflavodoxin from *Azotobacter vinelandii* with those available on all three other α–β parallel proteins the kinetic folding mechanism of which has been studied, i.e. *Anabaena* apoflavodoxin, *Fusarium solani pisi* cutinase and CheY. Two kinetic folding intermediates, one on-pathway and the other off-pathway, seem to be present during the folding of proteins with an α–β parallel, also called flavodoxin-like, topology. The on-pathway intermediate lies on a direct route from the unfolded to the native state of the protein involved. The off-pathway intermediate needs to unfold to allow the production of native protein. Available simulation data of the folding of CheY show the involvement of two intermediates with characteristics that resemble those of the two intermediates experimentally observed. Apparently, protein topology governs the appearance and kinetic roles of protein folding intermediates during the folding of proteins that have a flavodoxin-like fold.

Keywords: Protein folding; Topology; Flavodoxin; CheY; Cutinase; Folding intermediate

1. Introduction

Strong indications exist that the topology of a native protein influences the rate with which the native state is formed [1,2]. This implies that the transition state that determines the rate of protein folding resembles the native state of a protein. The influence of topology on folding rate is expressed in the contact order [3], which reflects the average sequence separation of residues that are in contact with one another in the native state. The larger the contact order, the slower a protein will fold.

Contact order can be used to predict the part of a protein that forms first during kinetic folding [4]. Parts of a protein that have a low local contact order, like e.g. α-helices, will form more rapidly than parts of a protein that have a high contact order, like e.g. parallel β-sheets. These predictions can in some cases be verified by comparison with experimentally determined Phi-values [5]. The Phi-value expresses to what extent native-like structure is formed around a specific residue in the transition state for folding, as determined from the effect a mutation has on the folding rate.

Although the simple contact order model has successfully captured the overall features of protein folding, it is shown to be an oversimplification of reality. Especially in larger proteins or proteins that display a certain degree of symmetry there are often several parallel routes to the native state that folding molecules can follow. These parallel routes may be equally consistent with the topological constraints on the transition state for folding and their population is determined by the magnitude of the local stabilities of the

* Corresponding author. Tel.: +31 317 484621; fax: +31 317 484801.
E-mail address: carlo.vanmierlo@wur.nl (C.P.M. van Mierlo).
[1] Present address: Department of Structural Biology, Faculty of Earth-and Life Sciences, Vrije Universiteit Amsterdam.

0301-4622/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.bpc.2004.12.005

structured protein regions formed in the corresponding transition states for folding. As a consequence, a mutation or a solvent condition that specifically stabilises part of a protein may significantly alter the observed folding mechanism. Incorporation of the free energy of the structures that can be formed on parallel routes into the contact order model considerably improves the ability of the model to predict structures of transition states [4].

A good approach to experimentally decipher the influence of the topology of a protein on its folding mechanism is by comparing the experimental results obtained on the folding of proteins which share the same topology but which have a low sequence homology. This has been done for the SH3 domains [6–8], acylphosphatase and procarboxypeptidase A2 [9,10], and immunoglobulin-like β-sandwich proteins [11]. Mutational analysis showed in these cases that indeed the transition state for folding is generally conserved among proteins with the same topology.

The question remains however, whether the appearance of folding intermediates and their specific role in kinetic folding is conserved among topologically related proteins that have no sequence homology. This aspect of protein folding has hardly been experimentally investigated so far [11,12]. The subject is difficult to resolve, because most protein folding intermediates form rapidly, often within the dead time of stopped-flow techniques, and are not substantially populated at equilibrium.

The study of the role of intermediates during folding is facilitated when these intermediates populate both kinetically and at equilibrium. The relatively large, 179-residue apoflavodoxin from *Azotobacter vinelandii* populates an intermediate with molten globule-like characteristics during denaturant- and thermally induced equilibrium unfolding [13,14]. In addition, *A. vinelandii* apoflavodoxin is shown to kinetically fold via two intermediates [15]. One of these intermediates is an off-pathway intermediate that populates heavily during apoflavodoxin kinetic refolding. It is shown that this intermediate populates during denaturant-induced equilibrium unfolding of apoflavodoxin [15]. The other intermediate is a high-energy folding intermediate that is only kinetically observed and lies on the direct folding route between unfolded and native apoflavodoxin.

Flavodoxins consist of a single structural domain and adopt the α–β parallel topology, also referred to as the doubly wound or flavodoxin-like topology. The flavodoxin-like topology is a rather popular fold: it belongs to the five most common observed folds, together with the TIM-barrel, Rossman, thiamin-binding and P-loop hydrolase folds [16]. In contrast to most protein folds, this α–β parallel topology is shared by many protein superfamilies (i.e. 16, according to the SCOP database [17], where the fold is referred to as "flavodoxin-like"). These superfamilies exhibit little or no sequence similarity and comprise a broad range of unrelated proteins with different functions like catalases, chemotactic proteins, lipases, esterases, and flavodoxins. They are all characterised by a five-stranded parallel β-sheet surrounded by α-helices at either side of the sheet (Fig. 1).

The PUBMED literature database was searched for kinetic folding data of the proteins categorised as flavodoxin-like in the SCOP database. Besides *A. vinelandii* apoflavodoxin, three proteins were identified the kinetic folding of which has been characterised in detail. Here, the folding kinetics of these four proteins are compared and the role protein topology has on the presence of intermediates during protein folding is investigated. Such a comparison leads to a better understanding of the fundamental rules describing the folding of proteins with a flavodoxin-like topology.

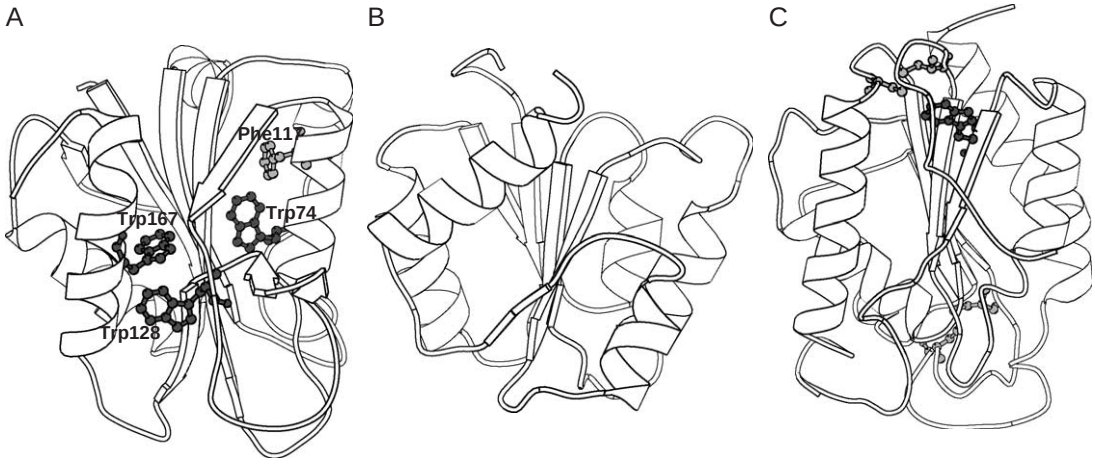


Fig. 1. Molscript cartoon drawings [28] of three proteins that share the α–β parallel, flavodoxin-like topology. (A) The X-ray structure of *Azotobacter chroococcum* flavodoxin [29], the sequence of which is 95% identical to *A. vinelandii* flavodoxin (B) *E. coli* CheY (3chy.pdb). (C) *F. solani pisi* cutinase (1agy.pdb). The side chains of the three tryptophan residues and of Phe117 of *A. vinelandii* flavodoxin are shown in ball-and-stick representation. In case of cutinase, its single tryptophan side chain and both disulfide bonds are shown.

2. Comparison of the folding of *A. vinelandii* apoflavodoxin with the folding of other flavodoxin-like proteins

2.1. A. vinelandii apoflavodoxin folding

The folding kinetics of *A. vinelandii* apoflavodoxin have been followed by stopped-flow experiments monitored by fluorescence intensity and anisotropy [15]. Single jump and interrupted refolding experiments show that the refolding kinetics involve four processes that all yield native molecules. Interrupted unfolding experiments show that the two slowest folding processes are due to Xaa-Pro peptide bond isomerisation in unfolded apoflavodoxin.

The denaturant dependence of the folding kinetics (i.e. the chevron plot) is complex. Under strongly unfolding conditions (>2.5 M GuHCl), single exponential unfolding kinetics is observed. The slope of the chevron plot changes between 3 and 5 M denaturant. This, together with the absence of an additional unfolding process reveals the presence of two consecutive transition states on a linear pathway that surround a high-energy on-pathway intermediate [15]. Under refolding conditions, two folding processes are observed for the folding of apoflavodoxin molecules with native Xaa-Pro peptide bond conformations, which implies the population of an intermediate. The slowest of these two processes becomes faster with increasing denaturant concentration, meaning that an unfolding step is rate-limiting for folding of the majority of apoflavodoxin molecules. This, together with the absence of a lag in the formation of native molecules, means that the intermediate that populates during refolding is off-pathway [15].

A molten globule-like intermediate I_1 populates during denaturant-induced equilibrium unfolding of apoflavodoxin. Its kinetic behaviour shows that it is not the high-energy on-pathway intermediate I_2 detected during apoflavodoxin unfolding. The experimental data obtained on apoflavodoxin folding are consistent with the linear four-state folding mechanism [15]:

$$I_1 \Leftrightarrow \text{unfolded} \Leftrightarrow I_2 \Leftrightarrow \text{native} \quad (1)$$

with the kinetic off-pathway intermediate I_1 being the one that also populates during denaturant-induced equilibrium unfolding of apoflavodoxin. I_1 has to unfold before native apoflavodoxin can be formed. All folding *A. vinelandii* apoflavodoxin molecules pass through I_2 before reaching the native state.

2.2. Comparison of the folding of A. vinelandii apoflavodoxin with CheY folding

The α–β parallel, flavodoxin-like protein the folding of which has been most intensely studied so far is CheY (129 amino acids, Fig. 1). CheY is both sequentially and functionally unrelated to flavodoxin. A structural alignment of 184 backbone atoms of CheY and of *A. vinelandii* apoflavodoxin with an rmsd of 1.62 Å shows 4% sequence identity and 20% similarity according to the PAM250 matrix [18]. The chevron plot recorded for CheY (un)folding displays a curvature on its folding side, indicative of the presence of an intermediate [19,20]. The CheY kinetic folding data are described by a three-state folding mechanism that involves an on-pathway intermediate I: U⇔I⇔N. However, according to Lopez-Hernandez et al. [20] it cannot be excluded that the observed intermediate is off-pathway because the first step in CheY folding is too fast to be resolved.

Quantitative analysis of kinetic folding data results in rate constants for folding and unfolding, and associated kinetic *m*-values. A kinetic *m*-value, i.e. the denaturant concentration dependence of a specific folding or unfolding rate constant, informs about the difference in denaturant accessible surface area between a specific folding species and the transition state involved [21]. In case a kinetic step implies an unfolding process, which is the case for example upon going from the native state to a partially folded transition state of a specific protein, the denaturant accessibility increases in this step and the corresponding *m*-value is positive. In case a kinetic step implies a folding process, which is the case for example upon going from the unfolded state to a partially folded transition state of a specific protein, the denaturant accessibility decreases in this step and the corresponding *m*-value is thus negative.

The kinetic *m*-values obtained for CheY folding contain indirect information about the role of the observed intermediate during kinetic folding of CheY. The trajectory the folding molecules follow in terms of free energy and solvent accessibility assuming the mechanism for CheY folding proposed by Lopez-Hernandez et al. [20] is valid is constructed by us and is shown in Fig. 2A.

As discussed, the mechanism used for the analysis of the experimental CheY folding data contains an on-pathway intermediate [20]. The use of this mechanism leads to *m*-values for the I to N and the N to I transition that are both positive (the reported values are both negative, due to the opposite definition of the *m*-value by Lopez-Hernandez et al. [20] compared to the one used here). This implies that the transition state between the intermediate and the native state is less compact than the intermediate, and that the on-pathway intermediate of CheY has to unfold partially before productive folding to the native state occurs. Although the latter is not impossible, it is unlikely. Instead, just as in case of *A. vinelandii* apoflavodoxin folding, we think that the observed curvature on the folding side of the chevron plot of CheY is better explained by an off-pathway intermediate.

The folding behaviour of a number of CheY mutants has been studied. Inspection of Fig. 1D in Lopez-Hernandez et al. [20] clearly shows that the kinetic refolding limb of the chevron plot of the Hel2 mutant of CheY, a mutant in which helix two has been stabilised, has a positive slope at low denaturant concentrations. A positive slope (i.e. a positive *m*-value) implies that an unfolding process is rate-limiting

in the formation of native CheY molecules. In our opinion, this positive slope confirms the involvement of an off-pathway folding intermediate during CheY folding. In addition, the observation made for CheY that stabilisation of any of its five helices causes slower refolding at low denaturant concentrations compared to native CheY supports our conclusion that this intermediate, in which these helices are structured, is off-pathway.

Furthermore, inspection of Fig. 2 in Munoz et al. [22] reveals a curvature in the unfolding limb of the chevron plot of wild-type CheY folding, similar to as is observed in the kinetic *A. vinelandii* apoflavodoxin unfolding data [15]. In more recent studies of CheY folding, this curvature is no

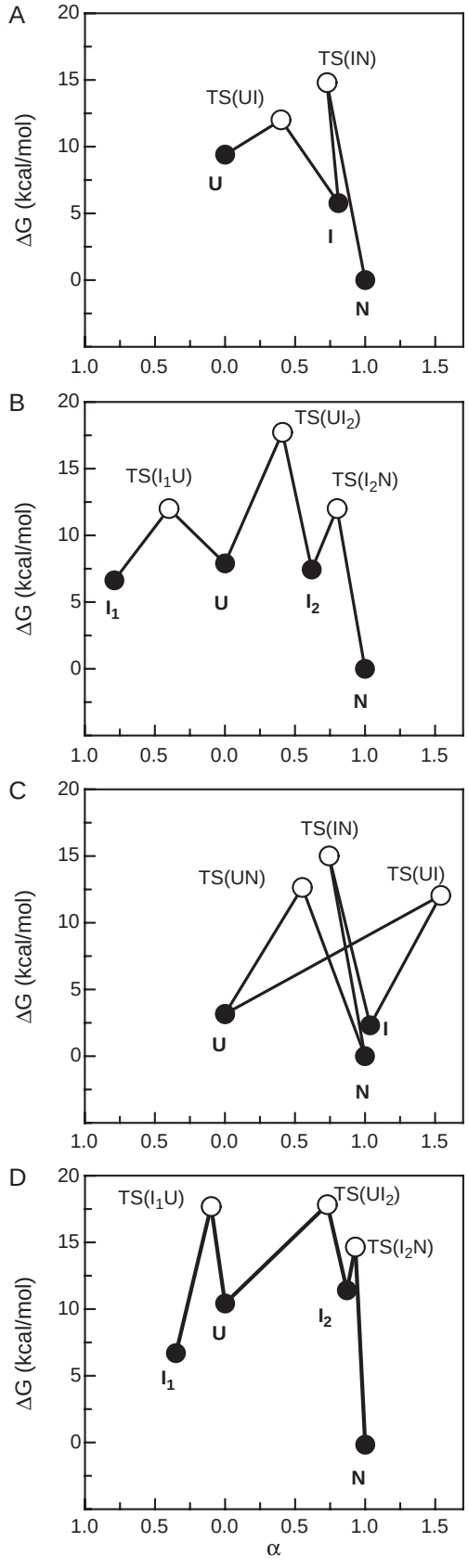


Fig. 2. Schematic representation of the folding trajectories of *E. coli* CheY (A), of *F. solani pisi* cutinase (B) and of apoflavodoxin from *Anabaena* (C) and from *A. vinelandii* (D). The free energies and denaturant accessibilities of the unfolded (U), intermediate (I, I_1, I_2), and native (N) states and of the corresponding transition states of the four proteins are shown on the vertical and horizontal axis, respectively. The denaturant-dependence of the free energy difference between a particular species and the unfolded state of the protein, i.e. its *m*-value, scales to the difference in denaturant-accessible protein surface between the particular species and the unfolded state. The denaturant accessibility of a folding species is expressed as the α-value. The α-value is the normalised *m*-value, i.e. unfolded protein has an α-value of 0 and native protein an α-value of 1. I represents the CheY or the *Anabaena* apoflavodoxin folding intermediate, I_1 and I_2 are the *A. vinelandii* apoflavodoxin folding intermediates presented in Scheme 1, and the corresponding intermediates in cutinase (see main text). TS(xy) is the transition state between species *x* and *y*. The heights of the barriers $\Delta G^{\ddagger}$ are calculated from the reported folding and unfolding rates according to $\Delta G^{\ddagger}=-RT\ln(k/k_0)$ using a value for k_0 of 10^8 [30]. Folding molecules follow the lines connecting two points. Note that in contrast to the proposed folding schemes for *Anabaena* apoflavodoxin and CheY folding, the denaturant accessibility of folding *A. vinelandii* apoflavodoxin molecules and of folding cutinase molecules gradually decrease on their productive folding route to the native state. (A) Folding trajectory of CheY constructed by us assuming that the protein folds according to the linear folding scheme proposed by Lopez-Hernandez et al. [19,20]: unfolded⇔intermediate⇔native. Note that the transition state between native CheY and its folding intermediate, TS(IN), is 10% less solvent accessible than the intermediate. The values for the free energy and denaturant accessibility of TS(UI) are chosen arbitrarily but in such a manner that they are consistent with the reported formation of I being too rapid to be observed. The data shown are extracted from Refs. [19,20] for the pseudo-wild-type protein (i.e. the F14N mutant of CheY). (B) Folding trajectory of *F. solani pisi* cutinase assuming that the protein folds according to the linear four-state folding scheme $I_1 \Leftrightarrow U \Leftrightarrow I_2 \Leftrightarrow N$ as proposed by Melo et al. [23]. The off-pathway intermediate I_1 is represented on the left-hand side of the unfolded state, whereas both the on-pathway intermediate I_2 and native cutinase reside on the right-hand side of the unfolded state. The transition states TS(I_1U) and TS(I_2N) are positioned arbitrarily, because cutinase folding from U to I_1 and cutinase unfolding from N to I_2 are too fast to be resolved [23]. (C) Folding trajectory of *Anabaena* apoflavodoxin constructed by us assuming that the protein folds according to the triangular folding Scheme 2 proposed by Fernández-Recio et al. [24] (see main text). Note that the transition state between U and I, TS(UI), is 50% less solvent accessible than native *Anabaena* apoflavodoxin. The α-values show that the transition between I and N involves an unfolding process. The free energy- and α-values presented are calculated based on the results of Fernández-Recio et al. [24]. (D) Folding trajectory of *A. vinelandii* apoflavodoxin assuming that the protein folds according to the linear four-state folding scheme $I_1 \Leftrightarrow U \Leftrightarrow I_2 \Leftrightarrow N$ as proposed by Bollen et al. [15] (see main text).

longer observed, which is probably due to the change of buffer from 5 mM sodium phosphate to 50 mM PIPES. The curvature in the unfolding limb of the chevron plot of CheY is most likely caused by the presence of a high-energy on-pathway intermediate on the direct (un)folding route of CheY.

In conclusion, the CheY folding data published suggest that kinetic on- and off-pathway intermediates similar to those observed during *A. vinelandii* apoflavodoxin folding are also involved during kinetic CheY folding.

2.3. Comparison of the folding of A. vinelandii apoflavodoxin with cutinase folding

Fusarium solani pisi cutinase is a lipolytic enzyme that adopts the α–β parallel, flavodoxin-like topology (Fig. 1). Despite their structural similarity, cutinase and *A. vinelandii* apoflavodoxin have a low sequence homology (14% identity and 40% similarity upon structural alignment). The folding and unfolding kinetics of cutinase have been studied by stopped-flow experiments [23]. The dependence of the folding and unfolding kinetics of cutinase on denaturant concentration (i.e. the chevron plot) strongly resembles the chevron plot of *A. vinelandii* apoflavodoxin: both the folding limb and the unfolding limb of the cutinase chevron plot are curved.

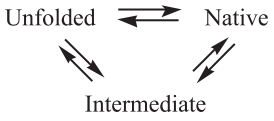
The curvature of the unfolding limb together with a change of the fluorescence intensity during the dead time of unfolding experiments indicate the transient population of an on-pathway intermediate when cutinase unfolds [23]. The curvature of the folding limb of the cutinase chevron plot is accompanied by a change of the fluorescence intensity during the dead time of the refolding experiments. This is interpreted as being caused by a second compact folding intermediate that is off the direct folding route of cutinase [23]. The latter conclusion is supported by the observation that in the presence of 0.75 M trehalose the folding limb of the cutinase chevron plot curves to such an extent that at low denaturant concentrations the rate with which native cutinase is formed increases with increasing denaturant concentration [23], just as is observed for *A. vinelandii* apoflavodoxin folding [15]. The increase of the folding rate upon increasing the concentration denaturant is characteristic for the involvement of an off-pathway intermediate during protein folding, because such an increase shows that an unfolding step is rate-limiting during the formation of native molecules. The trajectory the folding cutinase molecules follow in terms of free energy and solvent accessibility is shown in Fig. 2B.

In conclusion, also cutinase, just like *A. vinelandii* apoflavodoxin, folds via two intermediates: one is on the direct folding route and the other one is off the direct folding route that links unfolded and native cutinase. The observation of similar on- and off-pathway intermediates during the folding of CheY, cutinase and *A. vinelandii* apoflavodoxin suggests that apparently the appearance of these folding intermediates is governed by the flavodoxin-like topology these proteins share.

2.4. Comparison of the folding of A. vinelandii apoflavodoxin with Anabaena apoflavodoxin folding

Besides the presentation of the kinetic folding data of *A. vinelandii* apoflavodoxin by Bollen et al. [15], recently a rather limited number of kinetic folding experiments on an apoflavodoxin extracted from *Anabaena* have been reported [24]. In the latter study, a kinetic mechanism for *Anabaena* apoflavodoxin folding is proposed that involves only one intermediate that is off the direct route linking unfolded and native apoflavodoxin. The unfolding of this intermediate is rate-limiting at low urea concentrations, as is clear from the increase of one of the two observed *Anabaena* apoflavodoxin refolding rates upon increasing the denaturant concentration.

The kinetic *Anabaena* apoflavodoxin folding data are according to Fernandez-Recio et al. [24] best described by a three-state triangular model:



(2)

However, the folding intermediate of *Anabaena* apoflavodoxin as presented in Scheme 2 needs to have rather unusual properties, as recognised by Fernandez-Recio et al. Kinetic *m*-value analysis shows that its conformation seems to be as compact as the native state. In addition, the use of mechanism 2 for *Anabaena* apoflavodoxin folding leads to *m*-values for the I to N and the N to I interconversion via the common transition state TS(IN) that are both positive (i.e. m_{IN} is 0.48 ± 0.07 and m_{NI} is 0.42 ± 0.35 kcal/mol·M^{-1}, respectively [24]). This implies that the transition state between the intermediate and the native state is less compact than the intermediate, and that the on-pathway intermediate of *Anabaena* apoflavodoxin has to unfold partially before productive folding to the native state occurs.

Even more remarkable, the unfolding of the *Anabaena* apoflavodoxin folding intermediate to the unfolded state is associated with a negative *m*-value ($m_{IU}=-0.88\pm0.04$ kcal/mol·M^{-1} [24]). This negative *m*-value implies a peculiar, super-compact transition state to occur between unfolded *Anabaena* apoflavodoxin and its folding intermediate. This transition state is approximately 50% more compact than both the intermediate and the native state of the protein. Apoflavodoxin structures that are more compact than the native state might theoretically exist. Native apoflavodoxins are structurally for the largest part identical to holoflavodoxin except that they contain a relatively open, flexible flavin binding site [25], which potentially could close resulting in a transition state that is more compact than the native state is. However, the presence of an apoflavodoxin conformation that needs to be 50% more compact

than native apoflavodoxin to be able to explain the negative *m*-value discussed is highly unlikely.

The trajectory the folding apoflavodoxin molecules follow in terms of free energy and solvent accessibility, assuming the proposed triangular mechanism for *Anabaena* apoflavodoxin folding to be valid, is constructed by us and is shown in Fig. 2C. For comparison, in Fig. 2D the analogous trajectory for *A. vinelandii* apoflavodoxin folding is shown (data are taken from Ref. [15]). In case of cutinase and *A. vinelandii* apoflavodoxin folding the denaturant accessibility of the protein gradually decreases upon following the productive folding route that leads from unfolded to native protein (Fig. 2B, D). Instead, if *Anabaena* apoflavodoxin folding would occur according to Scheme 2, large unrealistic changes in solvent accessibility of the folding protein are observed on the different folding routes individual *Anabaena* apoflavodoxin molecules can follow (Fig. 2C).

In conclusion, the kinetic folding data published show that an off-pathway intermediate plays a significant role during *Anabaena* apoflavodoxin folding, just as is the case for *A. vinelandii* apoflavodoxin folding. However, the triangular *Anabaena* apoflavodoxin folding mechanism (Scheme 2) proposed by Fernandez-Recio et al. leads to, as discussed, some highly unlikely features of the protein states involved. Apparently, Scheme 2 does not correctly describe the kinetic folding of *Anabaena* apoflavodoxin. Further details of the kinetic mechanism of *Anabaena* apoflavodoxin folding might be revealed in the future by interrupted refolding and unfolding experiments similar to those presented by Bollen et al. [15] for *A. vinelandii* apoflavodoxin. Perhaps a change of buffer reveals that, just as in case of *A. vinelandii* apoflavodoxin, cutinase and CheY folding, besides an off-pathway intermediate an on-pathway intermediate plays a role as well in *Anabaena* apoflavodoxin folding.

2.5. Simulation results obtained for the folding of a flavodoxin-like protein

The folding of CheY has been studied theoretically using a simple model for protein folding by Clementi et al. [26]. The model consists of a Gō-like potential, in which the only forces used arise from contacts which are present in the native state, whereas the energetic frustration is drastically reduced by not including residue-specific parameters. As the latter contacts are highly conserved among topologically related proteins, the simulations by Clementi et al. are expected to inform about the influence topology has on folding. The simulations roughly reproduce the structure of the transition state for folding of CheY as determined experimentally by means of mutational analysis [19,20]. Furthermore, the simulations predict two possible intermediates: a short-lived on-pathway intermediate and an intermediate which is a misfolded trap [26]. The latter intermediate has all five helices which are present in native CheY well structured, but lacks its central parallel β-sheet. During the simulation, this misfolded species has to unfold, at least partially, before the native state is reached. Since non-native interactions are not possible in the model by Clementi et al. [26], the misfolded species is interpreted to be the result of topological constraints.

In conclusion, since the constraints in the theoretical model of Clementi et al. [26] are derived from contacts in the native state of CheY, application of this model to flavodoxin, cutinase or any other protein that shares the flavodoxin-like topology should produce the involvement of similar folding intermediates as theoretically predicted for CheY. An off-pathway intermediate is indeed observed in the experimental folding kinetics of all four α–β parallel proteins discussed. In addition, kinetic folding experiments show that the folding of three out of these four proteins, incidentally the three experimentally best characterised ones, involve an on-pathway intermediate.

3. Structural characteristics of the off-pathway intermediate observed during the folding of flavodoxin-like proteins

In this section, the structural characteristics of the off-pathway folding intermediates of *A. vinelandii* apoflavodoxin, CheY and cutinase are compared (unfortunately, no structural information is available on the intermediate observed during the folding of *Anabaena* apoflavodoxin). In case topology determines the formation of intermediates, the structures of these intermediates are expected to be similar.

The off-pathway intermediate I_1 of *A. vinelandii* apoflavodoxin (Scheme 1) populates during denaturant-induced equilibrium unfolding of the protein. A comparison of the fluorescence and CD data in the transition zone of unfolding leads to the conclusion that I_1 has conserved secondary structure (approximately 65% native α-helicity, see Ref. [15]) but that it lacks the characteristic tertiary structure of native apoflavodoxin. I_1 has molten-globule like features. This can explain the relatively low sensitivity of this state to denaturant (m_{UI_1}=−1.83 kcal/mol · M^{-1} versus m_{UN}=−6.23 kcal/mol · M^{-1} [15]). These features of I_1 are consistent with the observation by Clementi et al. [26] and by Lopez-Hernandez et al. [19,20] that the off-pathway intermediate in CheY folding has all native helices formed but lacks most of the long range interactions observed in the native state of the protein.

The off-pathway folding intermediate of CheY contains mainly helices as ordered structural elements, and ellipticity values indicate this to be the case for the corresponding folding intermediate of *A. vinelandii* apoflavodoxin as well. The helical nature of both off-pathway intermediates explains why they are formed so easily during CheY and apoflavodoxin folding. Helices are formed much more rapidly than sheets, especially when parallel β-sheets are

involved as is the case for apoflavodoxin and CheY. This 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 are separated by many residues from one another [27]. Apparently, rapid formation of α-helices followed by some docking prevents the formation of the parallel β-sheet of CheY and presumably that of apoflavodoxin.

The off-pathway intermediate observed during the folding of cutinase is relatively compact (α=0.79 compared to α=0.29 for the off-pathway intermediate of *A. vinelandii* apoflavodoxin; see the legend of Fig. 2 for an explanation of α-values). Cutinase contains a single Trp residue, the fluorescence of which is quenched in the native state by the vicinity of a disulfide bridge. Fluorescence data show that in the cutinase off-pathway intermediate the Trp fluorescence is not quenched despite that the aromatic residues of cutinase are buried from the solvent. Consequently, the native interaction between the Trp side chain and the disulfide bond that is adjacent in native cutinase is not formed in the cutinase off-pathway intermediate. This leads to the conclusion that the cutinase off-pathway intermediate contains non-native interactions [23].

Note that α–β parallel proteins contain two hydrophobic cores in their native state, which reside on the two opposite sides of the β-sheet and involve helices (Fig. 1). Non-native docking of these α-helices, without the β-sheet being formed or involved, in the off-pathway intermediates formed by α–β parallel proteins could be caused by hydrophobic interactions between the side chains of these helices. As a result the off-pathway folding intermediates have to unfold to yield native protein molecules.

In case of CheY folding, mutational analysis and simulations show that in the off-pathway intermediate all five α-helices are formed [20,26]. In the subsequent rate limiting transition state, helices 4 and 5 are unfolded, whereas helices 1 to 3 and β-strands 1 to 3 of native CheY are present [19,26]. Whether or not a similar situation occurs during *A. vinelandii* apoflavodoxin folding remains to be studied. However, the fluorescence anisotropy data presented by Bollen et al. [15] allow for some speculation about the conformation of the off-pathway intermediate. The native state of apoflavodoxin has a low anisotropy due to specific side-chain interactions as discussed by Bollen et al. [15]. The off-pathway apoflavodoxin folding intermediate, which also populates during equilibrium unfolding of the protein, has a fluorescence anisotropy that is higher than that of the native protein [15]. Consequently, the tryptophans that give rise to the fluorescence signal must be immobilised in the intermediate but the native interaction between Trp 128 and Trp 167 (Fig. 1) is not formed in the intermediate. The latter tryptophan residues are located in helix 5 and β-strand 5a of apoflavodoxin, respectively. Helix 5 and β-strand 5a may thus be formed in the off-pathway folding intermediate of *A. vinelandii* apoflavodoxin but they are definitely not natively oriented towards one another, as is also the case in the off-pathway intermediate of CheY [19,20,26].

In conclusion, the structural characteristics of the off-pathway folding intermediates of *A. vinelandii* apoflavodoxin, CheY and cutinase are similar. This is to be expected in case the conformation of off-pathway intermediates is determined to a substantial extent by the topology of the native state, which is the α–β parallel topology for the proteins discussed here. An off-pathway intermediate appears to be present in the folding of all four flavodoxin-like proteins the folding of which has been experimentally or theoretically investigated (i.e. CheY from *E. coli*, cutinase from *F. solani pisi* and apoflavodoxin from *A. vinelandii* and from *Anabaena*). It needs to unfold to allow the production of native protein.

Apparently, the constraints the flavodoxin-like topology puts on the polypeptide sequence result in two distinct minima in the free energy landscape for folding: one in which the native state resides, and one in which a relatively stable intermediate resides. Interconversion between these minima occurs via the globally unfolded state. This suggests that the intermediate, although it contains many native interactions, is effectively a misfolded species. Some interactions that stabilise the intermediate might be non-native and can be sequence-specific. This may explain the observed differences between the stabilities of the off-pathway intermediates found in the folding of different flavodoxin-like proteins. For example, the off-pathway folding intermediate of *A. vinelandii* apoflavodoxin populates during denaturant-induced equilibrium unfolding experiments, whereas the corresponding intermediates of CheY, cutinase and *Anabaena* apoflavodoxin do not, or hardly, populate at equilibrium.

4. Structural characteristics of the on-pathway folding intermediate observed during the folding of flavodoxin-like proteins

The on-pathway intermediate observed during the folding of *A. vinelandii* apoflavodoxin never significantly populates, neither kinetically nor at equilibrium. Consequently, it is not possible to extract the precise conformational properties of this intermediate from the apoflavodoxin folding data presented by Bollen et al. [15]. The on-pathway intermediate has a denaturant accessibility that is comparable to those of both transition states that surround it on the direct folding route of *A. vinelandii* apoflavodoxin (Fig. 2C). Consequently, the on-pathway folding intermediate probably structurally resembles these two transition states. The most denaturant-accessible one of these transition states (i.e. $TS(UI_2)$) is the one that limits the folding rate under native conditions [15]. The conformation of this transition state most likely resembles the one of the rate-limiting transition state of CheY folding because, as discussed, the structure of the rate-limiting

transition state is conserved among structurally homologous proteins.

In case of CheY folding, phi-value analysis shows that helices 1 to 3 and β-strands 1 to 3 are formed in the rate-limiting transition state, whereas the C-terminal part of the protein (helices 4 and 5 and strands 4 and 5) are unstructured [19,20]. The conformation of this transition state is roughly reproduced by folding simulations [26]. If the conformation of the on-pathway intermediate of α–β parallel proteins indeed resembles the one of the CheY transition state, its N-terminal half is structured, whereas its C-terminal half is unstructured.

The on-pathway intermediate of cutinase shares characteristics with the intermediate that populates during denaturant-induced equilibrium unfolding experiments of cutinase. Both species are around 40% unfolded in terms of solvent exposure [23]. Fluorescence data show that both in the on-pathway intermediate and in the equilibrium intermediate the environment of the single Trp side chain of cutinase is native-like. This Trp is located in β-strand 3. Thus, the conformation of the cutinase on-pathway intermediate might indeed resemble the conformation of the rate-limiting transition state of CheY, in which the N-terminal half of the protein is structured, whereas the C-terminal half of the conformation is unstructured.

5. Conclusion

The folding kinetics of four flavodoxin-like proteins have been compared. All four proteins appear to form an off-pathway intermediate during refolding. This intermediate needs to unfold before productive folding occurs. The structural details that are available about the off-pathway folding intermediates of *A. vinelandii* apoflavodoxin, of cutinase and of CheY are consistent with the off-pathway intermediate being a partially misfolded species. Many elements of native secondary structure appear to be formed in this species, but completion of productive folding to the native state is prohibited, probably by non-native docking of these elements onto one another.

Of the four proteins examined, *A. vinelandii* apoflavodoxin and *F. solani pisi* cutinase clearly form an on-pathway intermediate during their unfolding. In case of CheY unfolding, such an intermediate is seen when 5 mM sodium phosphate is used as buffer. In case of *Anabaena* apoflavodoxin unfolding, no on-pathway intermediate is observed. However, the kinetic data of *Anabaena* apoflavodoxin are not well described by the three-state kinetic folding model proposed [24], suggesting unresolved complexities. Perhaps a change in buffer conditions will reveal an on-pathway intermediate for this protein as well, just as is observed for CheY.

The on- and off-pathway intermediates experimentally observed during the folding of the α–β parallel proteins discussed resemble the two folding intermediates that are observed during the computational folding simulation of the α–β parallel protein CheY. Apparently, protein topology determines the appearance and kinetic roles (i.e. being on- or off-pathway) of protein folding intermediates during the folding of proteins that have a flavodoxin-like topology. This implies that the overall shape of the free energy landscape of protein folding depends on protein topology, whereas details like the depths of specific minima in this free energy landscape depend on peptide sequence.

Additional theoretical and experimental investigation of the folding of other proteins than the ones discussed here that share the α–β parallel topology is required. The results to be obtained for these topologically related proteins should contribute to a detailed understanding of the influence of the primary sequence on the structural and energetic characteristics of protein folding intermediates.

References

[1] D.P. Goldberg, Finding the right fold, Nat. Struct. Biol. 6 (1999) 987–990.
[2] D. Baker, A surprising simplicity to protein folding, Nature 405 (2000) 39–42.
[3] E. Alm, D. Baker, Prediction of protein-folding mechanisms from free-energy landscapes derived from native structures, Proc. Natl. Acad. Sci. U. S. A. 96 (1999) 11305–11310.
[4] V. Grantcharova, E.J. Alm, D. Baker, A.L. Horwich, Mechanisms of protein folding, Curr. Opin. Struct. Biol. 11 (2001) 70–82.
[5] A.R. Fersht, A. Matouschek, L. Serrano, The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding, J. Mol. Biol. 224 (1992) 771–782.
[6] D.S. Riddle, V.P. Grantcharova, J.V. Santiago, E. Alm, I. Ruczinski, D. Baker, Experiment and theory highlight role of native state topology in SH3 folding, Nat. Struct. Biol. 6 (1999) 1016–1024.
[7] J.C. Martinez, L. Serrano, The folding transition state between SH3 domains is conformationally restricted and evolutionarily conserved, Nat. Struct. Biol. 6 (1999) 1010–1016.
[8] K. Lindorff-Larsen, M. Vendruscolo, E. Paci, C.M. Dobson, Transition states for protein folding have native topologies despite high structural variability, Nat. Struct. Mol. Biol. 11 (2004) 443–449.
[9] V. Villegas, J.C. Martinez, F.X. Aviles, L. Serrano, Structure of the transition state in the folding process of human procarboxypeptidase A2 activation domain, J. Mol. Biol. 283 (1998) 1027–1036.
[10] F. Chiti, N. Taddei, P.M. White, M. Bucciantini, F. Magherini, M. Stefani, C.M. Dobson, Mutational analysis of acylphosphatase suggests the importance of topology and contact order in protein folding, Nat. Struct. Biol. 6 (1999) 1005–1009.
[11] J. Clarke, E. Cota, S.B. Fowler, S.J. Hamill, Folding studies of immunoglobulin-like beta-sandwich proteins suggest that they share a common folding pathway, Struct. Fold. Des. 7 (1999) 1145–1153.
[12] K. Gunasekaran, S.J. Eyles, A.T. Hagler, L.M. Gierasch, Keeping it in the family: folding studies of related proteins, Curr. Opin. Struct. Biol. 11 (2001) 83–93.
[13] C.P.M. van Mierlo, W.M. van Dongen, F. Vergeldt, W.J. van Berkel, E. Steensma, The equilibrium unfolding of *Azotobacter vinelandii* apoflavodoxin II occurs via a relatively stable folding intermediate, Protein Sci. 7 (1998) 2331–2344.
[14] C.P.M. van Mierlo, J.M. van den Oever, E. Steensma, Apoflavodoxin (un)folding followed at the residue level by NMR, Protein Sci. 9 (2000) 145–157.

[15] Y.J.M. Bollen, I.E. Sánchez, C.P.M. van Mierlo, Formation of on- and off-pathway intermediates in the folding kinetics of *Azotobacter vinelandii* apoflavodoxin, Biochemistry 43 (2004) 10475–10489.
[16] M. Gerstein, A structural census of genomes: comparing bacterial, eukaryotic, and archaeal genomes in terms of protein structure, J. Mol. Biol. 274 (1997) 562–576.
[17] A.G. Murzin, S.E. Brenner, T. Hubbard, C. Chothia, SCOP: a structural classification of proteins database for the investigation of sequences and structures, J. Mol. Biol. 247 (1995) 536–540.
[18] W.R. Pearson, Rapid and sensitive sequence comparison with FASTP and FASTA, Methods Enzymol. 183 (1990) 63–98.
[19] E. Lopez-Hernandez, L. Serrano, Structure of the transition state for folding of the 129 aa protein CheY resembles that of a smaller protein, CI-2, Fold. Des. 1 (1996) 43–55.
[20] E. Lopez-Hernandez, P. Cronet, L. Serrano, V. Munoz, Folding kinetics of Che Y mutants with enhanced native alpha-helix propensities, J. Mol. Biol. 266 (1997) 610–620.
[21] J.K. Myers, C.N. Pace, J.M. Scholtz, Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding, Protein Sci. 4 (1995) 2138–2148.
[22] V. Munoz, E.M. Lopez, M. Jager, L. Serrano, Kinetic characterization of the chemotactic protein from *Escherichia coli*, CheY. Kinetic analysis of the inverse hydrophobic effect, Biochemistry 33 (1994) 5858–5866.
[23] E.P. Melo, L. Chen, J.M. Cabral, P. Fojan, S.B. Petersen, D.E. Otzen, Trehalose favors a cutinase compact intermediate off-folding pathway, Biochemistry 42 (2003) 7611–7617.
[24] J. Fernandez-Recio, C.G. Genzor, J. Sancho, Apoflavodoxin folding mechanism: an alpha/beta protein with an essentially off-pathway intermediate, Biochemistry 40 (2001) 15234–15245.
[25] E. Steensma, C.P.M. van Mierlo, Structural characterisation of apoflavodoxin shows that the location of the stable nucleus differs among proteins with a flavodoxin-like topology, J. Mol. Biol. 282 (1998) 653–666.
[26] C. Clementi, H. Nymeyer, J.N. Onuchic, Topological and energetic factors: what determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? An investigation for small globular proteins, J. Mol. Biol. 298 (2000) 937–953.
[27] O. Bieri, T. Kiefhaber, Elementary steps in protein folding, Biol. Chem. 380 (1999) 923–929.
[28] P.J. Kraulis, MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures, J. Appl. Crystallogr. 24 (1991) 946–950.
[29] R.N.F. Thorneley, G.A. Ashby, M.H. Drummond, R.R. Eady, D.L. Hughes, G. Ford, P.M. Harrison, A. Shaw, R.L. Robson, J. Kazlauskaite, H.A.O. Hill, Flavodoxin and nitrogen fixation: structure, electrochemistry and posttranslational modification by coenzyme A, in: K. Yagi (Ed.), Flavins and Flavoproteins, 1993, Walter de Gruyter, Berlin, pp. 343–354.
[30] F. Krieger, B. Fierz, O. Bieri, M. Drewello, T. Kiefhaber, Dynamics of unfolded polypeptide chains as model for the earliest steps in protein folding, J. Mol. Biol. 332 (2003) 265–274.