Exploring the folding pathways of annexin I, a multidomain protein. II. Hierarchy in domain folding propensities may govern the folding process

ARTICLE *in* JOURNAL OF MOLECULAR BIOLOGY · JULY 1998

Impact Factor: 4.33 · DOI: 10.1006/jmbi.1998.1828 · Source: PubMed

5 AUTHORS, INCLUDING:

Françoise Ochsenbein
Atomic Energy and Alternative Energies Co…
32 PUBLICATIONS **811** CITATIONS
SEE PROFILE

Raphaël Guerois
Atomic Energy and Alternative Energies Co…
66 PUBLICATIONS **2,781** CITATIONS
SEE PROFILE

Article No. mb981828 *J. Mol. Biol.* (1998) **279**, 1177–1185

JMB

Exploring the Folding Pathways of Annexin I, a Multidomain Protein. II. Hierarchy in Domain Folding Propensities may Govern the Folding Process

Françoise Cordier-Ochsenbein[1], Raphaël Guerois[1] Françoise Russo-Marie[2], Jean-Michel Neumann[1] and Alain Sanson[1]*†

[1]*Département de Biologie Cellulaire et Moléculaire Section de Biophysique des Protéines et des Membranes and URA CNRS 2096, CEA Saclay, 91191 Gif sur Yvette Cedex, France*

[2]*Institut Cochin de Génétique Moléculaire, Unité INSERM U332, 22 rue Méchain 75014 Paris, France*

**Corresponding author*

In the context of exploring the relationship between sequence and folding pathways, the multi-domain proteins of the annexin family constitute very attractive models. They are constituted of four ~70-residue domains, named D1 to D4, with identical topologies but only limited sequence homology of approximately 30%. The domains are organized in a pseudochiral circular arrangement. Here, we report on the folding propensity of the D1 domain of annexin I obtained from overexpression in *Escherichia coli*. Unlike the D2 domain, which is only partially folded, the isolated D1 domain exhibits autonomous refolding in pure aqueous solution. Similarly, the D3 domain and D2-D3 module were obtained from expression in *E. coli* but were found to be largely unfolded. No conclusion could be drawn for the D4 domain because it was not possible to extract it from the bacterial inclusion bodies. The data allow us to propose a plausible scenario for the annexin I folding. This working model states that firstly the D1 domain folds, and the D2 and D3 domains remain partly unfolded, facilitating the docking of the D4 domain to the D1 domain. In a second step, the D1 and D4 domains dock, and D4 may fold if already not folded. The final step starts with the stabilization of the D1-D4 module. This stabilization is crucial for allowing the non-native local interactions inside the still partially unfolded D2 domain to switch to the native long-range interactions involving D4. This switch allows the complete folding of D2 and D3. The model proposes a sequential and hierarchical process for the folding of annexin I and emphasizes the role of both native framework and non-native structures in the process.

Keywords: protein folding; NMR; annexin; stability; non-native structures

Introduction

Understanding the process by which proteins reach their native structure is still a matter of deep and exciting research. While it is well known and clearly established that the three-dimensional structure of proteins is encoded in their sequence, the way these sequences also control the folding process is far from understood. The main reason for this is that there are very few proteins for which the folding process has been thoroughly analyzed as compared to the number of proteins for which the three-dimensional structure has been solved. A second important reason is probably that proteins for which the folding process has been thoroughly studied and better under-

†Also from Université P. et M. Curie, 9 Quai Saint-Bernard, Bât. C, 75005 Paris, France.

F.C.-O. and R.G. contributed equally to this work and should be considered as joint first authors.

Abbreviations used: COSY, correlated spectroscopy; DSS, 2,2-dimethyl-2-silapentane-5-sulfonic acid; FPLC, fast performance liquid chromatography; GdnHCl, guanidinium choride; GST, glutathione-S-transferase; HMQC, heteronuclear multiple-quantum correlation; HSQC, heteronuclear single-quantum correlation; NOE, nuclear Overhauser effect; NOESY, nuclear Overhauser enhancement spectroscopy; PCR, polymerase chain reaction; TFA, trifluoroacetic acid; TOCSY, total correlated spectroscopy.

0022–2836/98/251177–09 $25.00/0

stood are rather small, ranging from 50 to 150 residues; see, for instance (Dobson *et al.*, 1994; Fersht, 1995; Jennings & Wright, 1993; Miranker & Dobson, 1996).

There are two key concepts that may lead to a detailed understanding of how the sequence codes for the end-point of the folding process, i.e. the native structure, and for the intermediate states of this process: the concept of folding pathway and the concept of intrinsic structural propensity of a protein segment. We simply define here a folding pathway as an ensemble of successive, and even compulsory, conformational events that bring a protein from its unfolded state to its native structure. With this definition, we do not need here to refer to either the ''old'' or ''new'' view (Baldwin, 1995; Dill & Chan, 1997). Insofar as the acquisition of an approximately correct topology constitutes the bottleneck of the folding process, structural propensity is expected to play a crucial role in the earlier steps. Indeed, in the absence of the long-range tertiary interactions, the conformational ensemble of the protein is entirely governed by the intrinsic structural propensity of the different protein segments. This also means that, whenever the structural propensity is known in some detail for a protein, we are quite justified in referring to this knowledge in predicting the main aspects of the folding pathways of this protein.

To date, folding of few multi-domain proteins has been analyzed (Jaenicke, 1991; Mayr *et al.*, 1997; Minard *et al.*, 1989; Plaxco *et al.*, 1997). The intrinsic stability of the domains in these proteins was found to be quite variable. As an example, both domains of phosphoglycerate-kinase exhibit comparable stabilities and are able to fold independently (Minard *et al.*, 1989). Alternatively, homologous domains of fibronectin type III (Plaxco *et al.*, 1997) were found to present very different folding kinetics that coincide with their intrinsic stability variation.

In the context of exploring the sequence to folding pathway relationship, the multi-domain proteins of the annexin family constitute very attractive models because of their four ~70 residue domains, D1 to D4 (Figure 1), exhibiting an identical topology with only a limited sequence homology of approximately 30%. Domains interact with a hydrophobic interface to constitute two modules D1-D4 and D2-D3. The domain topology comprises five helix segments, named A to E, organized in a typical super-helix topology (Favier-Perron *et al.*, 1996; Huber *et al.*, 1992; Weng *et al.*, 1993). The advantage of these multi-domain proteins is that the domains of a given protein may have quite different stability and internal structural propensity in spite of their identical tertiary structure. Therefore, a detailed analysis of their secondary structure propensity as well as their tertiary structure propensity or stability may reveal important aspects of their folding pathway. Such an analysis may also reveal more general trends of the sequence to folding pathway relationship. Such work was therefore undertaken on annexin I.

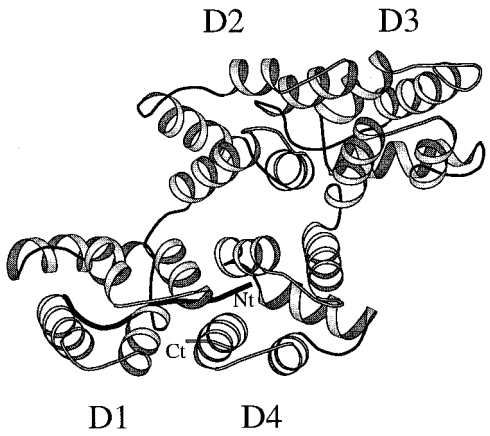

Figure 1. Crystallographic structure of annexin I designed using the program MOLSCRIPT (Kraulis, 1991).

In the accompanying paper (Cordier-Ochsenbein *et al.*, 1998), we have analyzed in detail the structural propensity within the D2 domain of annexin I and emphasized the role of non-native structures in stabilizing the unfolded state of this domain. Here, we report on the folding propensity of the other isolated domains and more specifically of the D1 domain obtained from expression in *Escherichia coli*. Unlike the D2 domain, the isolated D1 domain exhibits autonomous refolding in pure aqueous solution. In the same way, the D3 domain and D2-D3 module were obtained from expression in *Escherichia coli* and were found to be largely unfolded. No conclusion was drawn for the D4 domain, because it was not possible to extract this domain from the bacterial inclusion bodies. These data, combined with those presented in the accompanying paper, allowed us to propose a plausible scenario for the folding of annexins.

Results and Discussion

D1 domain is an autonomous folding unit

The uniformly ^{15}N-enriched D1 domain of annexin I was expressed in *E. coli* and purified as described in Materials and Methods. A preliminary analysis of the folding properties of the domain was undertaken using circular dichroism (CD) and multidimensional ^{1}H-^{15}N NMR.

The far-UV CD spectrum of the isolated D1 domain solubilized in pure aqueous solution is shown in Figure 2. As a first striking result, we observed that the D1 spectrum is characteristic of a large helix content, 50%, close to that found for the domain in the whole protein (~60%) and thus quite different from the 25% helix content exhibited by the isolated D2 domain. Furthermore, the associated near-UV CD spectrum (Figure 2(b)) indicates that the isolated D1 domain possesses a

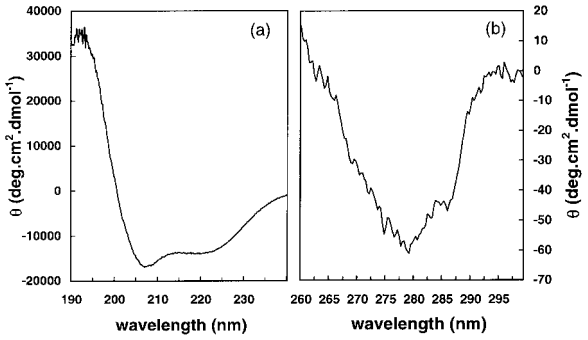

Figure 2. UV CD spectra of the isolated D1 domain recorded at 20°C: (a) far-UV spectrum (50 μM, acetate buffer, pH 3); (b) near-UV spectrum (10 μM, acetate buffer, pH 3).

tertiary structure. These data obviously reveal that two consecutive homologous domains of annexin I, D1 and D2, which adopt the same tertiary fold in the protein structure, exhibit a dramatic difference in stability when isolated from the rest of the protein.

Figure 3(a) shows part of the NOESY spectrum of D1 in aqueous solution exhibiting the NOE correlations arising from the amide and aromatic protons. For comparison, the corresponding spectral region obtained for D2 is reported (Figure 3(b)). The considerable difference in spectral dispersion observed between the D1 and D2 spectra is clearly consistent with the CD results. The D1 NOESY spectrum is characteristic of a folded polypeptide chain with, in particular, several chemical shifts close to 0 ppm and is quite different from that of the D2 domain exhibiting only residual helix structures (see Cordier-Ochsenbein *et al.*, 1998). Assignment of all the amide and Hα protons resonances as well as of a number of side-chain protons was obtained from standard 3D NOESY-HSQC, 2D

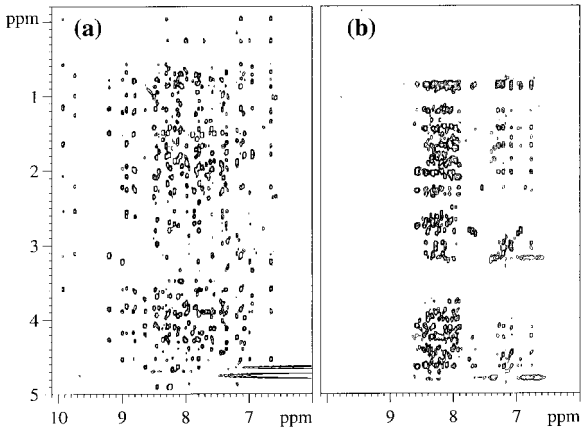

Figure 3. NOESY spectrum of the D1 and D2 isolated domains of annexin I in the region of cross-correlation between aliphatic and amide (and aromatic) protons. (a) D1 domain (4 mM, pH 3, 20°C); (b) D2 domain (4 mM, pH 6, 35°C).

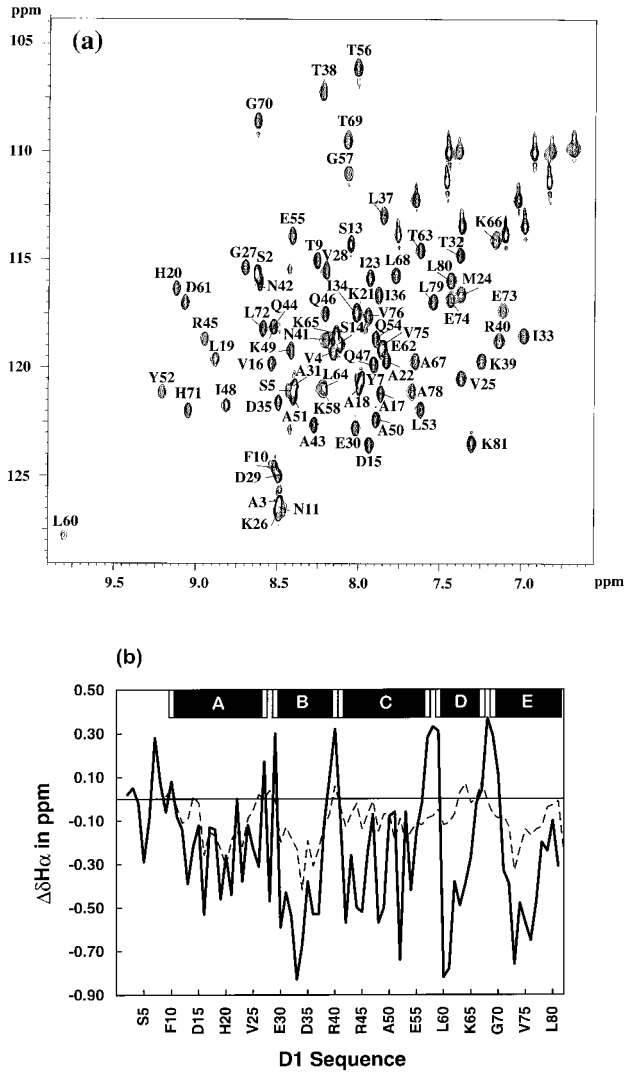

Figure 4. (a) $^{1}H/^{15}N$ HSQC spectrum of the D1 domain recorded at pH 3.0 and 20°C. Residue assignment is indicated. (b), Hα chemical shift index ΔδHα ($\delta H\alpha_{obs} - \delta H\alpha_{coil}$) obtained for the D1 domain solubilized in pure aqueous solution at pH 3 and 20°C (continuous line) and for the D2 domain solubilized in pure aqueous solution at pH 6 and 35°C (broken line). The Hα chemical shift coil values were taken from Merutka *et al.* (1995). Filled boxes on top of the Figure delineate the five native helices and the open boxes correspond to the transition residues.

HMQC-COSY, TOCSY and NOESY spectra. Figure 4(a) shows the $^{1}H/^{15}N$ HSQC spectrum of the uniformly ^{15}N-labelled D1 domain solubilized in pure aqueous solution at pH 3 recorded at 35°C. Since the D1 domain has its isoelectric point close to pH 6, a pH value of 3 was chosen to avoid aggregation. Figure 4(b) shows the profile drawn by the ΔδHα chemical shift indexes of D1 and, for comparison, that of D2. In agreement with the preceding results, the D1 domain exhibits ΔδHα amplitudes considerably larger than those of D2. Remarkably, this difference is observed for both

negative and positive values, i.e. concerns both helices and turn-like structures. The sequences of the five helices A to E found for the D1 domain in the crystallographic structure (Weng *et al.*, 1993) are reported in Figure 4(b). One can readily observe that they strictly coincide with those delineated by the $\Delta\delta H\alpha$ profile. All our data thus show that, in contrast with the D2 domain, the isolated D1 domain is able to adopt a native secondary structure and a tertiary structure most probably quite close to that of the domain in the intact protein.

Denaturation of the D1 domain by guanidinium chloride was followed by measuring the ellipticity at 222 nm (Figure 5(a)) or at 285 nm (Figure 5(b)). The curve exhibits the characteristic co-operative behavior of a folded protein with a transition midpoint above 2 M guanidinium chloride. Furthermore, for both far-UV and near-UV CD data, similar transition midpoints are found, indicating that secondary and tertiary structures unfold simultaneously on increasing the denaturant concentration. Using the fit procedure described by Pace (1986), the intrinsic free energy change ΔG_{buffer} and the associated *m*-value were estimated as 2.7(±0.3) kcal mol^{-1} and 1.2 kcal mol^{-1} mol^{-1} respectively. Interestingly, these values are close to

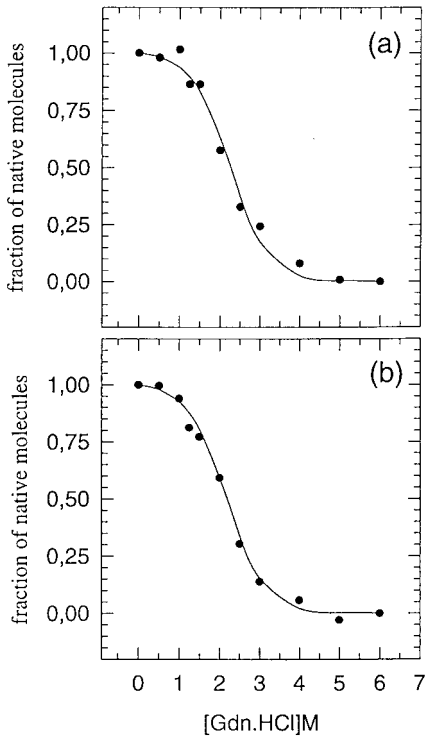

Figure 5. Guanidinium chloride-induced unfolding curve of the D1 domain, monitored by CD spectroscopy at 20°C: (a) at 222 nm and (b) at 285 nm. The experimental data were fit as described in Materials and Methods. The intrinsic free energy change ΔG_{buffer} and the associated *m*-value were estimated as 2.7(±0.3) kcal mol^{-1} and 1.2 kcal mol^{-1} mol^{-1}, respectively. The domain concentration was 10 mM in 20 mM acetate buffer, pH 3.

those observed for other isolated domains with equivalent sizes (Chen *et al.*, 1996; Viguera *et al.*, 1994) but are significantly lower than those found for small, single-domain proteins: CI2 (Jackson *et al.*, 1993), Sac7d (McCrary *et al.*, 1996), ubiquitin (Khorasanizadeh *et al.*, 1993) and barstar (Agashe & Udgaonkar, 1995)). This would suggest that the isolated D1 domain is not fully stabilized in the absence of interdomain interactions provided by the whole protein. However, a recent thermodynamic study on annexin V (Vogl *et al.*, 1997) showed that this protein exhibits a surprisingly low stability upon addition of GdnHCl ($\Delta G^{\circ}_{buffer}$ 2.5 kcal mol^{-1}). This low stability may correspond to a common feature of the annexin family and would rather indicate that the stability of the isolated D1 domain is close to that of this domain in the whole protein.

Comparison between the D1 and D2 domains of annexin I

The comparison between the D1 and D2 sequences readily provides key elements for understanding their striking difference in folding properties when isolated from the rest of the protein. First, the sequence homology between the two domains is rather low, i.e. ~30%, and essentially far from uniformly distributed within the five helix segments. Comparing helices, the sequence identity percentage decreases as follows: B-helix (58%), E-helix (36%), A-helix (30%), C-helix (23%) and D-helix (10%). Quite remarkably, as illustrated by the $\Delta\delta H\alpha$ profiles (Figure 4b) and precisely analyzed in the accompanying paper (Cordier-Ochsenbein *et al.*, 1998), this hierarchy is closely related to that of the helix propensity found in the residual secondary structure of the isolated D2 domain. The residual B-helix indeed appears as the most stable helix in the isolated D2 structure, while the most disordered part is found in the segment spanning the D-helix, which shares only one common residue with the corresponding motif in the D1 sequence.

The D1 and D2 domain sequences also differ when comparing the charged residues involved in electrostatic and hydrogen bond interactions stabilizing the non-native conformations of the isolated D2 domain and the corresponding amino acid residues in the D1 domain sequence. These differences are summarized in Figure 6. First, instead of the (FDxxEL) motif, a (SDxxAL) motif is found in the D1 sequence. The substitutions E (D2) to A (D1) and F (D2) to S (D1) prevent the formation of a non-native capping box in the A helix of D1. The second point concerns the K (D2) to G (D1) substitution at the C terminus of the C-helix. The Gly residue in D1 clearly acts as a helix stop signal. Conversely, Lys in D2 is not a helix-breaker residue although, in the native structure, this residue acts as a helix C-cap. In addition, the Lys residue is part of an ensemble of non-native interactions stabilizing the partially folded state of D2. Lastly, at

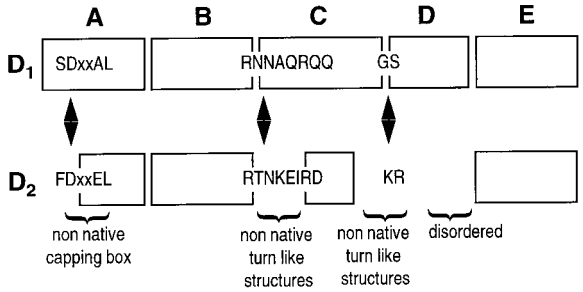

Figure 6. Comparison between the D1 and D2 sequences. (a) The five helices observed both in the annexin I crystallographic structure and the isolated domain are indicated as boxes. (b) The residual helix structures of the isolated D2 domain are indicated as boxes and non-native structures are explicitly indicated. The essential residues involved in the local non-native structures in the isolated D2 domain and their corresponding residues in the D1 domain sequence are reported.

the N terminus of the D2 C-helix, the RTNKEIRD segment that was found to adopt non-native local structures, stabilized by the various possible salt-bridges, is replaced in the D1 sequence by RNNAQRQQ, which includes no acidic residue (instead of two in the D2 sequence). Taken together, these observations show that all the residues involved in the stabilization of the non-native conformations detected in the isolated D2 domain correspond to different residue types in the D1 domain sequence. These sequence differences corroborate the idea proposed in the accompanying paper that the non-native conformations stabilize the partly folded state of the isolated D2 domain at the expense of the native state.

Isolated D3 domain and D2-D3 module of annexin I do not fold into a stable 3D structure

Purification of the D3 domain of annexin I was impossible because of extensive degradation by proteases occurring during its expression in bacteria. Similarly, we were unable to purify the D2-D3 module, since the full-length module was never detected in the crude extracts. Similar difficulties were encountered during purification of the isolated D2 domain, but there was enough intact domain for further purification (see Materials and Methods in the paper by Cordier-Ochsenbein *et al.*, 1998). These results strongly suggest (Hubbard *et al.*, 1994) that, as D2, D3 is only partially folded when isolated from the rest of the protein. As regards the D4 domain, we unfortunately cannot obtain any information on its intrinsic stability since, when expressed in *E. coli*, it remains in inclusion bodies and we have not yet been able to purify it. In summary, our results indicate a striking difference in the conformational properties of the four domains constituting annexin I. We will now discuss these results in the context of protein folding.

A scenario for the annexin folding process

Several general scenarios are readily imaginable for the folding process of a multi-domain protein. For annexins, the four domains could simply have had the same folding propensity or stability. If the four individual domains had low intrinsic stability, the folding of the whole protein could occur in a unique co-operative step. Alternatively, the four domains could be intrinsically stable, constitute autonomous folding units and thus fold independently. In the latter case the last step in the folding process of the annexin would be the closing of the pseudo-chiral circular arrangement of the four folded domains. In the case where the domains have different intrinsic stabilities, several other scenarios can be proposed such as, for instance, a scenario in which the D2-D3 module would constitute an autonomous folding unit able to bring D1 and D4 together (a general scenario proposed by Fink, 1995). Clearly, none of these scenarios fits our experimental data. Therefore, we suggest another model, which takes into account both experimental data and indirect evidence from the annexin structure.

Our working model corresponds to a sequential process and accounts for: (i) the marked hierarchy in structural propensities observed at the level of domain; (ii) the particular circular topology of annexins; and (iii) the existence of a well-identified conformational switch in the D2 domain as described in the accompanying paper. The sequence of events that we propose as building a main ensemble of folding pathways is depicted in Figure 7. It comprises three principal steps. (1) Experimental data show that the D1 domain constitutes an independent folding unit. This could constitute a general feature of the annexin family, since we found that the isolated D1 domain of annexin V also adopts a tertiary fold when isolated from the rest of the protein (data not shown). In contrast, the D2 and D3 domains as well as the D2-D3 module exhibit a rather low propensity to fold. This suggests that during the folding of annexin I, the D1 domain folds first, whereas the D2 and D3 domains fold in a later step of the folding process. The general hypothesis stating that the most stable structural elements of the protein architecture fold earlier during the folding process than the less stable ones is reinforced by several kinetic and equilibrium studies on apomyoglobin (Hughson *et al.*, 1990; Waltho *et al.*, 1993), lysozyme (Dobson *et al.*, 1994; Yang *et al.*, 1995), SNase (Shortle *et al.*, 1996), RNase H (Raschke & Marqusee, 1997) and fibronectin type III (Plaxco *et al.*, 1997). Incidentally, preliminary results (not shown) indicate that the isolated D1 domain of annexin I folds at least ten time faster than the entire annexin V protein. (2) Little information is available concerning the folding propensity of the

D4 domain. Nevertheless, several observations can be made that may help determine the fundamentals of the folding process of annexin I. First, it is obvious that the D4 domain must dock to the D1 domain at some stage of the folding process in order to establish the hydrophobic interface leading to the formation of the D1-D4 module (Figure 1) and to the final pseudo-chiral circular arrangement of the four domains. Closing the circular structure of the annexin constitutes a topological step that may be crucial for the folding process because it allows (i) the specific interactions between D2 and D4 domains to take place, (ii) the N and C-terminal parts of the protein to be close together. Second, examination of known annexin structures (Benz *et al.*, 1996; Bewley *et al.*, 1993; Favier-Perron *et al.*, 1996; Huber *et al.*, 1990, 1992; Weng *et al.*, 1993) reveals highly conserved specific interactions between the N-terminal tail and the D4 domain. Indeed, the N-terminal tail is inserted between the ends of the B and E-helices of D4, holding them together. This N-terminal tail is not part of the D1 domain, does not interfere with the stability of the D1 domain and is rather unstructured in the absence of D4 (data not shown) but could alternatively participate in the stability of D4. Interestingly, studies of proteins whose N and C-terminal parts are close in the native structure showed that the interaction and the stabilization of the N and C-terminal parts are important, and may even occur in the early steps of the folding process: cytochrome *c* (Kuroda *et al.*, 1995) and RNase H (Kanaya & Kanaya, 1995). Docking of D4 to the early-formed D1 native structure may proceed from a diffusive process involving random searchs. (3) This search will be all the more facilitated by the fact that parts of the protein in the D2-D3 region remain flexible when D4 is not docked to D1. This confers an important role on the non-native structures observed in the isolated D2 domain. These non-native structures contribute to maintain the D2 domain in a partially folded state until the D4 domain is sufficiently folded on contact with D1 to offer more favorable non-local interactions to the residues precisely involved in the non-native structures. This step corresponds to the conformational switch described in the accompanying paper, which leads to the folding of D2 and D3. The conformational switch mainly involves the residues of the D2 domain which, in association with several residues of D4, constitute the hydrophylic central core of the folded annexin.

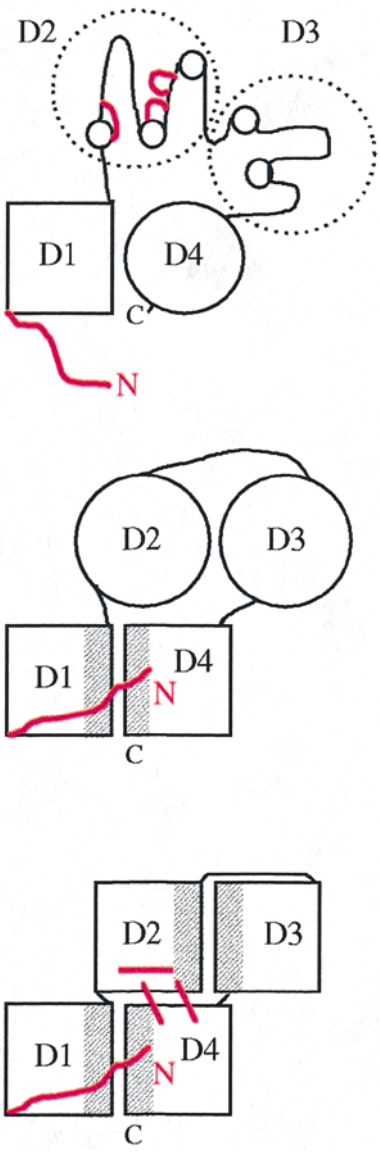

Figure 7. Scenario for the annexin I folding process. This working model states that D1 domain folds first, D2 and D3 domains are maintained partly unfolded by local non-native interactions (red loops) to allow D4 domain to dock to D1 domain. In a second step, D1 and D4 domains are docked, D4 may fold if already not folded. A first hydrophobic interface corresponding to the D1-D4 domain is created (hatched region) and the N-terminal tail can interact with D4 domain. As a final step, the stabilization of the D1-D4 module allows the switch from non-native local interactions to native long-range interactions (red sticks), and the folding of D2 and D3 domains. The second hydrophobic interface is created corresponding to the D2-D3 module.

The folding model that we propose for annexin I emphasizes the role of the framework elements (Ptitsyn, 1987). The ensemble of frameworks that were detected comprises at least one autonomous folding unit, the full D1 domain, and the segments of D2 with high helix propensity. Other frameworks are certainly present in the D3 and D4 domains. These may act, at different stages, as pre-positioned structures contributing to the co-operativity of the folding process. Interestingly, the highest helix-propensity segments of the D2 domain correspond to the A, B and E-helices, whose fate is to interact with the other domains in the whole protein (Cordier-Ochsenbein *et al.*, 1996). The present model also emphasizes the possible role of the different kinds of non-native structures and of the segments with low intrinsic

folding propensity in promoting sequentiality in the folding process by their kinetic influence (Hamada *et al*., 1995, 1996; Kuroda *et al*., 1996; Roder & Colon, 1997).

Material and Methods

Protein purification

Production and purification of the uniformly ^{15}N-enriched D1 domain. The D1 domain was produced by overexpression in *E. coli*. The cDNA of annexin I D1 domain was prepared using PCR from the cDNA provided by Biogen Incorp. The cDNA was inserted in the pGEX-2 T vector (Smith & Johnson, 1988). The absence of PCR-induced mutations was checked by gene sequencing. The sequence of the recombinant domain is the following: $\underline{G_1SAVS_5PYPTF_{10}}NPSSD_{15}VAALH_{20}KAIMV_{25}$-$KGVDE_{30}ATIID_{35}ILTKR_{40}NNAQR_{45}QQIKA_{50}AYL$-$QE_{55}TGKPL_{60}DETLK_{65}KALTG_{70}HLEEV_{75}VLALL_{80}K$. The underlined segment corresponds to a part of the N terminal tail of annexin I. Peptide production was achieved by growth of *E. coli* (BL21) containing the expression vector in minimal medium (4 g/l glucose, 0.66 g/l $(^{15}NH_4)_2SO_4$, 3 g/l KH_2PO_4, 14 g/l Na_2HPO_4, 0.5 g/l NaCl, 0.47 g/l $MgSO_4$, 5 mg/l thiamine and trace elements) at 37°C. After induction with isopropylthiogalactopyranoside (100 μM IPTG) at an absorbance of 1 at 600 nm, the growth was continued until the plateau was reached (about three hours). After centrifugation, bacteria were resuspended in the lysis buffer (50 mM Tris-HCl (pH 8), 10 mM EDTA, 500 mM NaCl, 5% (v/v) glycerol, 1% (v/v) Triton X100, 1 mM dithiothreitol (DTT), 1 mM phenylmethylsulfonyl fluoride (PMSF) and 20 μg/ml aprotinin). Purification was achieved as follows. After sonification and centrifugation (10,000 ***g***), the supernatant of soluble proteins was incubated with glutathione/agarose beads, allowing specific binding of the GST-D1 domain fusion protein to the beads. After washing (with 1 M NaCl, 50 mM Tris-HCl, pH 8), thrombin (70 units per liter of culture) was added and the domain eluted. The D1 domain was further purified on a proRPC (16/10) column (Pharmacia, France) using the FPLC system with a linear gradient of Millipore water (containing 0.1% (v/v) TFA) and acetonitrile (containing 0.1% TFA) at a flow-rate of 2.5 ml/minute. After solvent evaporation, the lyophilized domain was resuspended in NMR buffer. The final yield of purification was about 8 mg of domain per liter of culture.

CD experiments

CD spectra were recorded on a Jobin-Yvon (Longjumeau, France) CD6 spectrodichrograph, using a 0.5 mm path-length cell for far-UV spectra and a 1 cm path-length cell for near-UV spectra. The emission spectral bandwidth, the wavelength increment and the accumulation time were 2 nm, 0.2 nm and 0.8 second per step, respectively. Far and near-UV spectra were obtained by averaging three and 20 successive spectra, respectively. The solvent spectrum was recorded under identical conditions and subtracted. The secondary structure content was estimated according to the molar ellipticity at 222 nm assuming a θ value of 31,020 deg cm^2 $dmol^{-1}$ for 100% helix content. The peptide concentration was 10 mM. The solvent used was 20 mM acetate buffer (pH 3). The temperature was kept at 298 K.

NMR experiments

A portion (17.6 mg) of uniformly ^{15}N-labeled annexin I D1 domain obtained as described above was solubilized in 0.5 ml (4 mM) of 20 mM phosphate buffer, 0.1 mM EDTA, 2 mM NaN_3, 0.1 mM DSS and 10% ^2H_2O. The pH was adjusted to 3.0. NMR spectra were recorded on a Bruker DRX 600 equipped with a *z*-gradient accessory. Standard homonuclear COSY, TOCSY, NOESY (Markley, 1989; Wüthrich, 1986) and HSQC (Lerner & Bax, 1986), gradient-HSQC, gradient 3D NOESY-HSQC (Grzesiek & Bax, 1993), HMQC-COSY (Gronenborn *et al*., 1989) experiments were performed. Either a MLEV 17 (Bax & Davis, 1985) or a DIPSI 2 pulse scheme (Brown & Sanctuary, 1991) with 80 ms mixing time was used for TOCSY spectra. Mixing times varying from 75 to 200 ms were used for NOESY spectra. A GARP ^{15}N-decoupling sequence was used in all the experiments using uniformly ^{15}N-labeled domain. Water suppression was achieved using either the presaturation, jump and return sequence (Plateau & Guéron, 1982) or the WATERGATE sequence (Piotto *et al*., 1992). Spectra were recorded at different temperatures, ranging from 278 K to 313 K. In general, spectra were acquired with a recycling delay of one second. The 2D homonuclear spectra were recorded with 400 increments of 2 K data points, 2D heteronuclear spectra with 256 increments of 2 K data points and 3D experiments with 200 increments in the proton dimension and 80 increments in the ^{15}N dimension of 512 data points. Shifted sine-bell functions were used for apodization. The 2D and 3D data were processed using XWINNMR software (Bruker) and TRIAD software (Tripos), respectively. ^1H chemical shifts were referenced with DSS. ^{15}N chemical shifts were referenced according to the $^1H/^{15}N$ frequency ratio of 0.101329118, as reported by Wishart *et al*. (1995).

Acknowledgments

This work was partly supported by a fellowship (R.G.) from the Institut de Formation Supérieure Biomédicale.

References

Agashe, V. R. & Udgaonkar, J. B. (1995). Thermodynamics of denaturation of barstar: evidence for cold denaturation and evaluation of the interaction with guanidine hydrochloride. *Biochemistry*, **34**, 3286–99.

Baldwin, R. L. (1995). The nature of protein folding pathways: the classical versus the new view. *J. Biomol. NMR*, **5**, 103–109.

Bax, A. & Davis, D. G. (1985). MLEV-17-based two dimensional homonuclear magnetisation transfer spectroscopy. *J. Magn. Reson.* **65**, 355–360.

Benz, J., Bergner, A., Hofmann, A., Demange, P., Gottig, P., Liemann, S., Huber, R. & Voges, D. (1996). The structure of recombinant human annexin VI in crystals and membrane- bound. *J. Mol. Biol.* **260**, 638–643.

Bewley, M. C., Boustead, C. M., Walker, J. H., Waller, D. A. & Huber, R. (1993). Structure of chicken annexin V at 2.25-Å resolution. *Biochemistry*, **32**, 3923–3929.

Brown, L. R. & Sanctuary, B. C. (1991). Hetero-TOCSY experiments with WALTZ and DIPSI mixing sequences. *J. Magn. Reson.* **91**, 413–421.

Chen, Y. J., Lin, S. C., Tzeng, S. R., Patel, H. V., Lyu, P. C. & Cheng, J. W. (1996). Stability and folding of the SH3 domain of Bruton's tyrosine kinase. *Proteins: Struct. Funct. Genet.* **26**, 465–471.

Cordier-Ochsenbein, F., Guerois, R., Baleux, F., HuynhDinh, T., Chaffotte, A., Neumann, J. M. & Sanson, A. (1996). Folding properties of an annexin I domain: A H-1–N-15 NMR and CD study. *Biochemistry,* **35**, 10347–10357.

Cordier-Ochsentein, F., Guerois, R., Baleux, F., Huynh-Dinh, T., Lirsac, P.-N., Russo-Marie, F., Neumann, J.-M. & Sanson, A. (1998). Exploring the folding pathways of annexin I, a multi-domain protein. I. Non-native structures stabilize the partially folded state of the isolated domain 2 of annexin I. *J. Mol. Biol.,* **279**, 1163–1175.

Dill, K. A. & Chan, H. S. (1997). From Levinthal to pathways to funnels. *Nature Struct. Biol.* **4**, 10–19.

Dobson, C. M., Evans, P. A. & Radford, S. E. (1994). Understanding how proteins fold: the lysozyme story so far. *Trends Biochem. Sci.* **19**, 31–37.

Favier-Perron, B., Lewit-Bentley, A. & Russo-Marie, F. (1996). The high-resolution crystal structure of human annexin III shows subtle differences with annexin V. *Biochemistry,* **35**, 1740–1744.

Fersht, A. R. (1995). Optimization of rates of protein folding: the nucleation-condensation mechanism and its implications. *Proc. Natl Acad. Sci. USA,* **92**, 10869–10873.

Fink, A. L. (1995). Compact intermediates states in protein folding. *Subcell. Biochem.* **24**, 27–53.

Gronenborn, A. M., Bax, A., Wingfield, P. T. & Clore, G. M. (1989). A powerful method of sequential proton resonance assignment in proteins using relayed 15N-1H multiple quantum coherence spectroscopy. *FEBS Letters,* **243**, 93–98.

Grzesiek, S. & Bax, A. (1993). The importance of not saturing H_2O in protein NMR. Application to sensitivity enhancement and NOE measurements. *J. Am. Chem. Soc.* **115**, 12593–12594.

Hamada, D., Kuroda, Y., Tanaka, T. & Goto, Y. (1995). High helical propensity of the peptide fragments derived from beta-lactoglobulin, a predominantly beta-sheet protein. *J. Mol. Biol.* **254**, 737–746.

Hamada, D., Segawa, S. & Goto, Y. (1996). Non-native alpha-helical intermediate in the refolding of beta-lactoglobulin, a predominantly beta-sheet protein. *Nature Struct. Biol.* **3**, 868–873.

Hubbard, S. J., Eisenmenger, F. & Thornton, J. M. (1994). Modeling studies of the change in conformation required for cleavage of limited proteolytic sites. *Protein Sci.* **3**, 757–768.

Huber, R., Romisch, J. & Paques, E. P. (1990). The crystal and molecular structure of human annexin V, an anticoagulant protein that binds to calcium and membranes. *EMBO J.* **9**, 3867–3874.

Huber, R., Berendes, R., Burger, A., Schneider, M., Karshikov, A., Luecke, H., Romisch, J. & Paques, E. (1992). Crystal and molecular structure of human annexin V after refinement. Implications for structure, membrane binding and ion channel formation of the annexin family of proteins. *J. Mol. Biol.* **223**, 683–704.

Hughson, F. M., Wright, P. E. & Baldwin, R. L. (1990). Structural characterization of a partly folded apomyoglobin intermediate. *Science,* **249**, 1544–1548.

Jackson, S. E., Moracci, M., elMasry, N., Johnson, C. M. & Fersht, A. R. (1993). Effect of cavity-creating mutations in the hydrophobic core of chymotrypsin inhibitor 2. *Biochemistry,* **32**, 11259–11269.

Jaenicke, R. (1991). Protein folding: local structures, domains, subunits, and assemblies. *Biochemistry,* **30**, 3147–3161.

Jennings, P. A. & Wright, P. E. (1993). Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin. *Science,* **262**, 892–896.

Kanaya, E. & Kanaya, S. (1995). Reconstitution of *Escherichia coli* RNase HI from the N-fragment with high helicity and the C-fragment with a disordered structure. *J. Biol. Chem.* **270**, 19853–19860.

Khorasanizadeh, S., Peters, I. D., Butt, T. R. & Roder, H. (1993). Folding and stability of a tryptophan-containing mutant of ubiquitin. *Biochemistry,* **32**, 7054–7063.

Kraulis, P. J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946–950.

Kuroda, Y., Endo, S., Nagayama, K. & Wada, A. (1995). Stability of alpha-helices in a molten globule state of cytochrome c by hydrogen-deuterium exchange and two-dimensional NMR spectroscopy. *J. Mol. Biol.* **247**, 682–688.

Kuroda, Y., Hamada, D., Tanaka, T. & Goto, Y. (1996). High helicity of peptide fragments corresponding to β-strand regions of β-lactoglobulin observed by 2D-NMR spectroscopy. *Fold. Des.* **1**, 255–263.

Lerner, L. & Bax, A. (1986). Sensitivity enhanced two dimentional heteronuclear relayed coherence transfer NMR spectroscopy. *J. Magn. Reson.* **69**, 375–385.

Markley, J. L. (1989). Two-dimensional nuclear magnetic resonance spectroscopy of proteins: an overview. *Methods Enzymol.* **176**, 12–64.

Mayr, E. M., Jaenicke, R. & Glockshuber, R. (1997). The domains in gammaB-crystallin: identical fold-different stabilities. *J. Mol. Biol.* **269**, 260–269.

McCrary, B. S., Edmondson, S. P. & Shriver, J. W. (1996). Hyperthermophile protein folding thermodynamics: differential scanning calorimetry and chemical denaturation of Sac7d. *J. Mol. Biol.* **264**, 784–805.

Merutka, G., Dyson, H. J. & Wright, P. E. (1995). 'Random coil' 1H chemical shifts obtained as a function of temperature and trifluoroethanol concentration for the peptide series GGXGG. *J. Biomol. NMR,* **5**, 14–24.

Minard, P., Hall, L., Betton, J. M., Missiakas, D. & Yon, J. M. (1989). Efficient expression and characterization of isolated structural domains of yeast phosphoglycerate kinase generated by site-directed mutagenesis. *Protein Eng.* **3**, 55–60.

Miranker, A. D. & Dobson, C. M. (1996). Collapse and cooperativity in protein folding. *Curr. Opin. Struct. Biol.* **6**, 31–42.

Pace, C. N. (1986). Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods Enzymol.* **131**, 266–280.

Piotto, M., Saudek, V. & Sklenar, V. (1992). Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J. Biomol. NMR,* **2**, 661–665.

Plateau, P. & Guéron, M. (1982). Exchangeable proton NMR without base-line distortion, using new strong-pulse sequence. *J. Am. Chem. Soc.* **104**, 7310–7311.

Plaxco, K. W., Spitzfaden, C., Campbell, I. D. & Dobson, C. M. (1997). A comparison of the folding kinetics and thermodynamics of two homologous fibronectin type III modules. *J. Mol. Biol.* **270**, 763–770.

Ptitsyn, O. B. (1987). Protein folding: hypotheses and experiments. *J. Protein Chem.* **6**, 273–293.

Raschke, T. M. & Marqusee, S. (1997). The kinetic folding intermediate of ribonuclease H resembles the acid molten globule and partially unfolded molecules detected under native conditions [published erratum appears in Nature Struct. Biol. 1997 Jun;4(6):505]. *Nature Struct. Biol.* **4**, 298–304.

Roder, H. & Colon, W. (1997). Kinetic role of early intermediates in protein folding. *Curr. Opin. Struct. Biol.* **7**, 15–28.

Shortle, D., Wang, Y., Gillespie, J. R. & Wrabl, J. O. (1996). Protein folding for realists: a timeless phenomenon. *Protein Sci.* **5**, 991–1000.

Smith, D. B. & Johnson, K. S. (1988). Single-step purification of polypeptides expressed in *Escherichia Coli* as fusions with glutathione-*S*-transferase. *Gene,* **67**, 31–40.

Viguera, A. R., Martinez, J. C., Filimonov, V. V., Mateo, P. L. & Serrano, L. (1994). Thermodynamic and kinetic analysis of the SH3 domain of spectrin shows a two-state folding transition. *Biochemistry,* **33**, 2142–2150.

Vogl, T., Jatzke, C., Hinz, H. J., Benz, J. & Huber, R. (1997). Thermodynamic stability of annexin V E17G: equilibrium parameters from an irreversible unfolding reaction. *Biochemistry,* **36**, 1657–1668.

Waltho, J. P., Feher, V. A., Merutka, G., Dyson, H. J. & Wright, P. E. (1993). Peptide models of protein folding initiation sites. 1. Secondary structure formation by peptides corresponding to the G- and H-helices of myoglobin. *Biochemistry,* **32**, 6337–6347.

Weng, X., Luecke, H., Song, I. S., Kang, D. S., Kim, S. H. & Huber, R. (1993). Crystal structure of human annexin I at 2.5 Å resolution. *Protein Sci.* **2**, 448–458.

Wishart, D. S., Bigam, C. G., Yao, J., Abildgaard, F., Dyson, H. J., Oldfield, E., Markley, J. L. & Sykes, B. D. (1995). 1H, 13C and 15N chemical shift referencing in biomolecular NMR. *J. Biomol. NMR,* **6**, 135–140.

Wüthrich, K. (1986). *NMR of Proteins and Nucleic Acids*, John Wiley and Sons, New York.

Yang, J. J., Buck, M., Pitkeathly, M., Kotik, M., Haynie, D. T., Dobson, C. M. & Radford, S. E. (1995). Conformational properties of four peptides spanning the sequence of hen lysozyme. *J. Mol. Biol.* **252**, 483–491.

Edited by A. R. Fersht

(Received 4 November 1997; received in revised form 25 March 1998; accepted 27 March 1998)