

Protein Unfolding Transitions in an Intrinsically Unstable Annexin Domain: Molecular Dynamics Simulation and Comparison with Nuclear Magnetic Resonance Data

Tru Huynh,* Jeremy C. Smith,*† and Alain Sanson*

*Commissariat à l'Energie Atomique-Saclay, Département de Biologie Joliot-Curie/Service de Biophysique des Fonctions Membranaires and Unité de Recherche Associée Centre National de la Recherche Scientifique 2096, 91191 Gif-sur-Yvette Cedex, France; and

†Lehrstuhl für Biocomputing, Interdisziplinäres Zentrum für Wissenschaftliches Rechnen, Universität Heidelberg, Im Neuenheimer Feld 368, 69120 Heidelberg, Germany

ABSTRACT Unfolding transitions of an intrinsically unstable annexin domain and the unfolded state structure have been examined using multiple ~ 10 -ns molecular dynamics simulations. Three main basins are observed in the configurational space: native-like state, compact partially unfolded or intermediate compact state, and the unfolded state. In the native-like state fluctuations are observed that are nonproductive for unfolding. During these fluctuations, after an initial loss of $\sim 20\%$ of the core residue native contacts, the core of the protein transiently completely refolds to the native state. The transition from the native-like basin to the partially unfolded compact state involves $\sim 75\%$ loss of native contacts but little change in the radius of gyration or core hydration properties. The intermediate state adopts for part of the time in one of the trajectories a novel highly compact salt-bridge stabilized structure that can be identified as a conformational trap. The intermediate-to-unfolded state transition is characterized by a large increase in the radius of gyration. After an initial relaxation the unfolded state recovers a native-like topology of the domain. The simulated unfolded state ensemble reproduces in detail experimental nuclear magnetic resonance data and leads to a convincing complete picture of the unfolded domain.

INTRODUCTION

A microscopic description, at atomic detail, is required for a complete understanding of protein folding and unfolding. All-atom molecular dynamics (MD) simulation is in principle a powerful tool for obtaining such a detailed view. However, although progress is promising (Ferrara and Caffisch, 2000; Daura et al., 1999; Duan and Kollman, 1998), MD simulation of protein folding remains beyond available computer power. In contrast, unfolding simulations can provide useful information (e.g., Prompers et al., 2001; Mayor et al., 2000; Inuzuka and Lazaridis, 2000; Wong et al., 2000; Gilquin et al., 2000; Tsai et al., 1999; Smith et al., 1999; Tirado-Rives et al., 1997; Lazaridis and Karplus, 1997; Lamy and Smith, 1996). Although one can question to what extent present unfolding simulations provide useful information on the folding process (Finkelstein, 1997; Daura et al., 1999), it is clear that as timescales accessible lengthen with increasing available computer power, unfolding simulations will approach equilibrium conditions so as to provide a description of the different states of the system accessed in reversible folding processes.

The proteins of the annexin family constitute particularly attractive models for studying folding. Their four ~ 70 residue domains, D1 to D4 (Fig. 1 *a*), exhibit similar tertiary topology but only a limited sequence homology of $\sim 30\%$. The domain structure comprises five helix segments, A to E (Fig. 1 *b*), organized in a super-helix topology (Favier-Perron et al., 1996; Huber et al., 1992; Weng et al., 1993). The strong structural symmetry of the annexin topology masks a strongly asymmetric amino acid sequence and, as a consequence, a strong asymmetry of structural propensities of fragments and variations in domain stability. For example, although isolated domain 1 is an autonomous folding unit, domains 2 and 3 are intrinsically unfolded (Cordier-Ochsenbein et al., 1996, 1998a).

Domain 2 is a particularly interesting model for MD as it exists in an equilibrium unfolded state without denaturant. Moreover, the unfolded state of domain 2 has been extensively characterized using nuclear magnetic resonance (NMR) spectroscopy (Cordier-Ochsenbein et al., 1996, 1998a). Many details concerning residual native and non-native local structures, including precise helix content, was obtained but no long-range information such as the average topology of the unfolded domain could be deduced. Information on residual topology in the unfolded state is indeed important for understanding the folding process of a multidomain protein such as annexins with the idea that sequence may not only code for the final three-dimensional structure but also for some key events of the folding process. Such an information, which is usually inaccessible to experiment, can be obtained from simulation of molecular dynamics, provided a critical level of agreement with experimental data is reached.

Submitted November 25, 2001, and accepted for publication February 28, 2002.

Tru Huynh's present address is Institut Pasteur - Bioinformatique Structurale, 25-28 rue du Docteur Roux, 75724 Paris, Cedex 15 France.

Alain Sanson is also from Université Pierre et Marie Curie, Paris, France.

Address reprint requests to Alain Sanson, CEA-Saclay, DBJC/SBFM and URA CNRS 2096, 91191 Gif-sur-Yvette Cedex, France. Tel.: 33-1-69-08-28-63; Fax: 33-1-69-08-81-39; E-mail: sanson@dsvidf.cea.fr.

© 2002 by the Biophysical Society

0006-3495/02/08/681/18 \$2.00

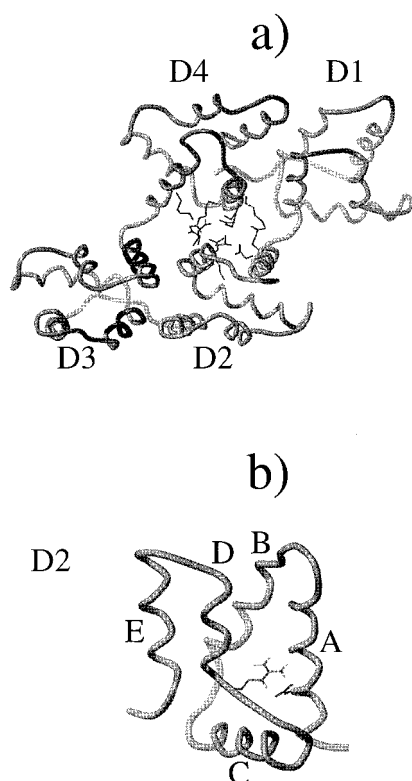


FIGURE 1 Structure of annexin I (a) and of its domain 2 (b). The conserved salt-bridge between A and B helices is also indicated in b, together with the hydrophilic interactions between domain 2 and 4 in a; darker helices in domain 3 are the partners of B and E helices of domain 2 in the hydrophobic D2-D3 interface.

For the present MD experiment the nonequilibrium process of excision of the domain from the native structure followed by unfolding was examined as follows:



During the unfolding simulations a number of interesting phenomena were observed. First, during the first ~ 3 ns the native core unfolds then spontaneously refolds completely to the native state before again unfolding. A partially unfolded compact state is subsequently formed, which for part of the time, in one trajectory, is a highly salt-bridge stabilized structure as compact as the native state. Finally, structural changes in the unfolded state are found to be in detailed agreement with the NMR data.

MATERIALS AND METHODS

Model system and potential function

The molecular dynamics simulations were performed with CHARMM (Brooks et al., 1983), version c26b2, using the PARM22 all-atom parameter

set (MacKerrel et al., 1998). The starting structure was that determined by x-ray crystallography (Weng et al., 1993; S. Kim, personal communication.).

Various simulation protocols and methods for treating long-range interactions were tested using preliminary simulations (not shown) on the A-helix fragment in aqueous solution, for which a well-defined NMR structure exists (Guerois et al., 1998). Based on these simulation results, the method finally selected was to truncate the electrostatic interactions with atom-based force switching over a range of 8 to 12 Å, and to treat the Van der Waals forces with the shift method and a 12-Å cutoff. Moreover, this truncation scheme has been shown to be efficient and accurate in previous work (Boczko and Brooks, 1995; Chatfield et al., 1998; Steinbach and Brooks, 1994) and also yielded good agreement between experimental and simulation data in simulations of annexin 1 domain 1 (Huynh et al., 1999).

The domain 2 was explicitly solvated with TIP3P water molecules (Jorgensen et al., 1983). For this the domain was inserted in a preequilibrated cubic water box. Water molecules whose oxygen atoms were closer than 2.8 Å to any heavy atom of the solute were removed. As the net charge of domain 2 is -3 , to ensure the system neutrality three sodium ions were manually added by replacing three water molecules. Two box sizes were used, one of 62.04 Å side containing initially 8000 water molecules giving a final system size of 23,565 atoms, and one for the unfolded state simulation of 80 Å side containing initially 17,576 water molecules with a final size of 50,163 atoms. Periodic boundary conditions were applied to the solvent and ions, centered around the center of mass of the protein domain.

In comparison with previous work on unfolding simulations the present work uses more extensive solvation and a large nonbonded cutoff value (12 Å). The box size was chosen so as to avoid interactions between proteins in the main and image-boxes, which were found to lead to unfolding artifacts (T. Huynh and A. Sanson, unpublished data). One consequence of these simulation parameter is that they lead to physically more realistic nonexplosive protein unfolding (i.e., see Huynh et al., 1999). However, the computer time required per step and the number of steps required for unfolding are then significantly greater. Consequently it is impossible to perform a large number of simulations with present computer power and the results consequently have reduced statistical significance. Hence, most of the unfolding transition behavior seen here gives only a qualitative indication of possible equilibrium unfolding transitions.

Simulation protocol

To efficiently solvate the protein the following protocol was used, during which the protein atoms were fixed: 1) two cycles of minimization were performed each with 2000 steps of the steepest descent algorithm; 2) the water was heated to 50 K, equilibrated for 5 ps, and cooled to 20 K; 3) two further cycles of 2000-step steepest descent minimization were performed; 4) the water molecules were fixed and the protein subjected to two cycles of 1000-step steepest descent energy minimization; and 5) all constraints were removed and the complete system subjected to two final cycles of 2000-step steepest descent minimization.

The leap frog algorithm was used to integrate the equations of motion in the MD simulations with a 1-fs time-step. The nonbonded lists were truncated at 14 Å. Periodic boundary conditions were applied to the solvent and ions around the center of mass of the protein domain and updated every 10 ps (10,000 steps). Heating was performed by increasing the temperature by 1 K every 100 steps. Subsequently, equilibration was performed in two stages of 50 ps each. During the first stage, every 50 steps the velocities

were assigned from a Gaussian distribution. For the second stage, the velocities were scaled if the temperature was outside a 10-K window around the desired temperature. After equilibration, the production phase was performed in the NVT ensemble. Six simulations were performed, one each at 300, 350, and 400 K and three at 450 K. The six trajectories are labeled in the text by their temperature. The three 450-K trajectories are further distinguished as 450K-A, 450K-B, and 450K-U as described in the text. Although the isolated domain has experimentally been shown to be unstable in aqueous solution at 300 K, the time-scale accessible by MD simulation is not long enough for unfolding to be observed at this temperature. Consequently, raised temperatures were needed, and the lower temperature simulations (300, 350, and 400 K) were used primarily to explore the native state conformational fluctuations.

Most of the computing power for the simulations was provided by the Cray T3E of Center d'Etudes Nucléaire de Grenoble, the IBM SP2SC of the CINES (Montpellier) and a 16-node Beowulf PC cluster. The 62-Å side simulations required 14,400 CPU days on the Cray T3E (225 days \times 64 CPUs). The larger solvation box system required 1648 CPU days on the PC cluster (103 days \times 16 CPUs).

RESULTS

Description of the system

The annexin I structure and that of the domain 2 is shown in Fig. 1, *a* and *b* respectively. The conserved salt-bridge between A and B helices is also indicated. In the full protein, domain 2 interacts with domain 3 via the hydrophobic interface BE(do2)-EB(do3) and with domain 4 via the hydrophilic interface AB(do2)-BA(do4). Stabilization of the latter interface during the compaction process of the full protein, is supposed to allow the folding of domain 2 and that of domain 3 (Cordier-Ochsenbein et al., 1998b). The domain 2 sequence is given below where the five helices are underlined and the core residues are indicated with bold italic characters:

GTPAQFDADE₁₀LRAAMKGLGT₂₀DEDTLIEILA₃₀SRT
NKEIRDI40NRVYREELKR50DLAKDITSDT₆₀SGDFRN
ALLS₇₀LAKG.

Description of the different simulations

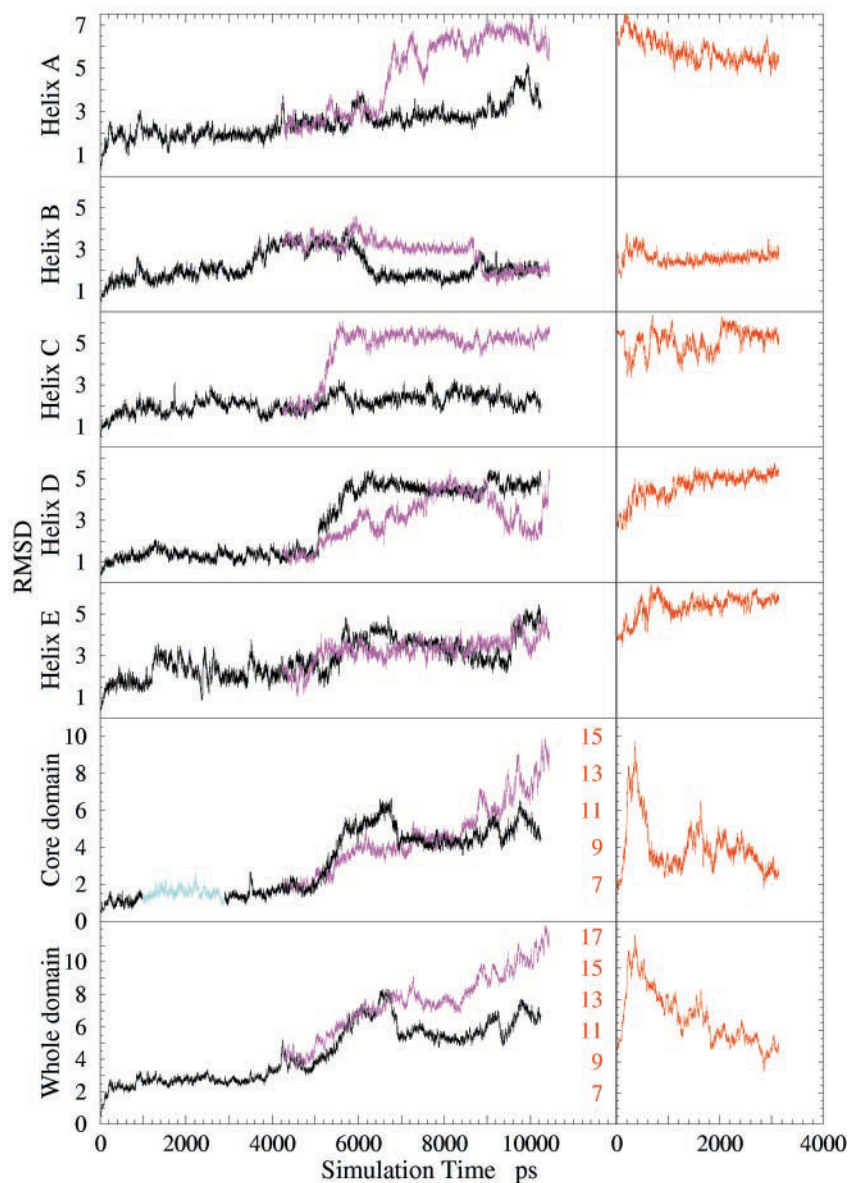
Although the isolated domain has experimentally been shown to be unstable in aqueous solution at 300 K, the time-scale accessible by MD simulation is not long enough for unfolding to be observed at this temperature. Consequently, raised temperatures were needed. Therefore, six simulations were performed, one each at 300, 350, and 400 K and three at 450 K. The lower temperature simulations (300, 350, and 400 K) were run for exploring the native state conformational fluctuations and the initial unfolding steps. The different trajectories are labeled in the text by their temperature. Two effective \sim 10-ns unfolding trajectories were run at 450 K and labeled 450K-A, 450K-B. A complementary 450 K, named 450K-U, was run for comparison with NMR data.

The three 450-K simulations will now be described. The first 10-ns trajectory performed, starting from the crystal

structure is named 450K-A. After 10 ns, the structure remained compact and conserved most of its native secondary structure, in contrast with the NMR results on the unfolded state, which indicate more extensive secondary structure modification. Therefore, it was decided to run a new unfolding simulation. However, to save computer time, instead of starting from the initial folded state, the new simulation was started at \sim 4.2 ns, thus creating a bifurcation of the initial trajectory. The 4.2-ns time point was chosen because a structural fluctuation occurs there involving breaking and quick reforming of the conserved D₇-R₃₂ salt-bridge between the A and B helices. This structural fluctuation is suspected to be an important interhelical stabilizing interaction in the domain, and so it was decided to perform the second bifurcating simulation (450K-B) starting from the frame of maximal salt-bridge separation. This second simulation immediately diverged from the first and the salt-bridge did not subsequently reform.

At the end of 450K-B, the annexin domain had unfolded more than 450K-A with an increased radius of gyration but had still not reached the equilibrium unfolded state as indicated by the NMR data and still contained more residual secondary structure than observed experimentally. Moreover, the size of the unfolding protein was too large for the initial solvent box size. To obtain a realistic model for the unfolded protein it was therefore decided to place the system in a larger simulation box and to subject the protein to an additional perturbation over a short period so as to reduce the extent of the helix-helix interactions while initially preserving the internal helix structure. The subsequent relaxation period would then provide information on intrinsic helix stability, i.e., in the absence of tertiary structural stabilization. The use of certain methods for applying this perturbation, such as a higher simulation temperature or radius of gyration constraint, would run the risk of perturbing the internal helix structure. Instead, here a less biasing, softer perturbation method was used. Starting from the 9635-ps frame of run 450K-B, weak constraints were applied to repel the helices from each other in such a way that no substantial secondary structure distortion could occur: 14 repulsive constraints (10 Kcal/mol/Å²) were applied (A/B, B/C, B/D, and E/D). However, in initial trials large distortions of helix structure were nevertheless observed, due to side-chain: side-chain salt-bridges. It was therefore decided to prevent salt-bridges from forming by also applying side-chain to side-chain constraints on 15 residues (50 Kcal/mol/Å²). During this procedure harmonic position constraints were also applied to the C-helix backbone to prevent distortion of this helix during application of the constraints on the other helices. Constraints were applied until no significant contact between helices A, B, C, and E was observed, i.e., for 145 ps. Using this method, the goal was achieved of reducing the extent of the helix-helix interactions without significantly altering the secondary structure. At this point, all the constraints were removed and the domain left to relax for \sim 3 ns (the 450K-U trajectory).

FIGURE 2 RMS deviation from minimized crystal structure of the domain 2 and its four helices. (Black) 450K-A trajectory; (magenta) 450K-B trajectory; (red) 450K-U trajectory. Fluctuation period in the native basin is colored in cyan.



The unfolded state: validation of the simulation by comparison with NMR data

To assess the quality of the simulation of nonequilibrium states such as the unfolding trajectories of proteins, the details of which are globally inaccessible to experiment, it is very important that simulations reproduce experimental results on equilibrium states. Numerous data on the equilibrium-unfolded state of the annexin I domain 2 have been obtained by NMR (Cordier-Ochsenbein et al., 1996, 1998a), and we thus start by describing the unfolded state in the 450K-U simulation obtained as indicated above.

During the 450K-U simulation the domain performs large amplitude fluctuations and relaxes to a structure that is more compact than the initial 450K-U configuration (Fig. 2, whole domain root-mean-square deviation (RMSD)). Re-

markably, the native topology, which was considerably distorted after the force-unfolding step, is practically recovered after ~ 2 ns (Fig. 3). In addition, during the relaxation of the global structure, the amount of secondary structure decreases.

We now analyze the secondary structure and compare it with the NMR-derived data. Fig. 4a presents the NMR data (^1H chemical shift indices) for annexin I domain 2 in pure aqueous and micellar solutions (Cordier-Ochsenbein et al., 1998a). For the annexin domain, the micellar environment has the effect of enhancing the population of the local structures without changing their nature (Cordier-Ochsenbein et al., 1998a). This is true for both the native residual structures and, more importantly in the present case, for the nonnative local structures, which can thus be determined

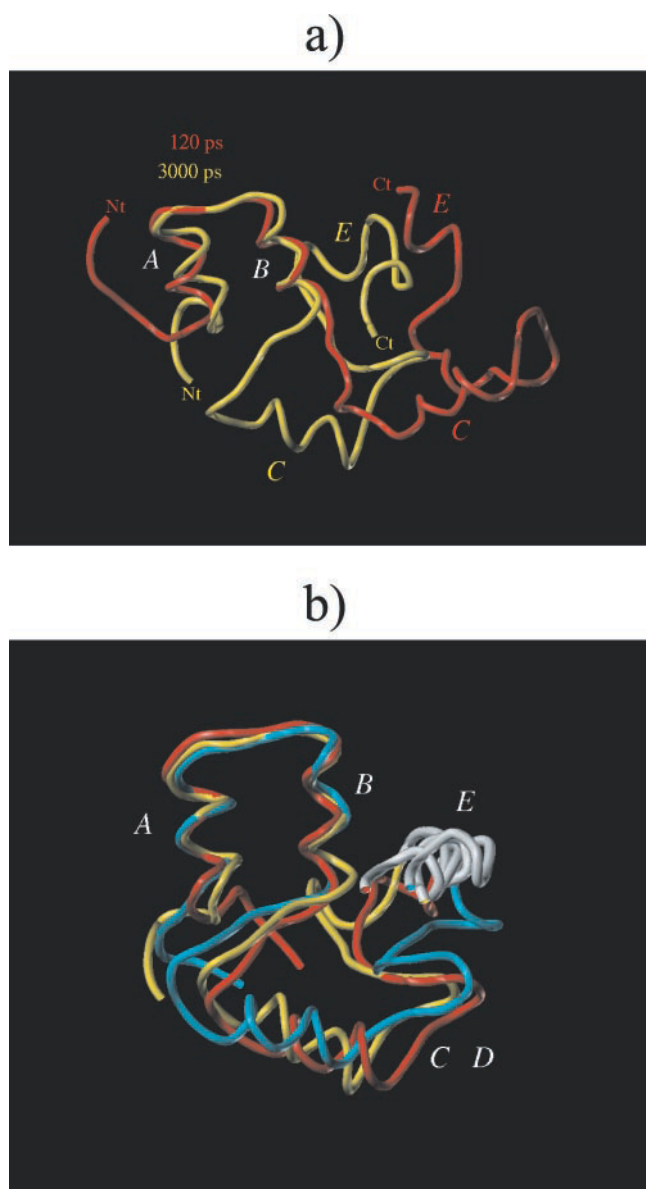


FIGURE 3 (a) Large scale movement of C, D, and E helices during the structure relaxation in the 450K-U trajectory. (b) Snapshots of the domain structure in the unfolded state (450K-U, last 500 ps).

with better precision and confidence. The heavy black line in Fig. 4 *a* clearly demonstrates this effect.

The chemical shift indices are positive when a residue is in the β region of the Ramachandran plot and negative when in the α region. The absence of a characteristic near-ultraviolet circular dichroism spectrum for the protein indicates the absence of significantly stable tertiary structure in the protein. However, in Fig. 4 *a* most of the residues have a negative chemical shift index, indicating that a significant amount of residual helix structure is present.

The experiments indicate that the helix population is concentrated in the B-helix (close to 100%), in the A₈-L₁₈ part of the A-helix and, to a lesser extent, in the E-helix. The

C and D helices behave as a single segment with only a small helix population in the central part of the C-helix.

It is difficult to calculate ¹H chemical shift indices directly from simulation-derived structural ensembles as the index depends not only on ϕ - ψ values but also on other environmental properties the effects of which are difficult to calculate. ¹³C chemical shift indices are in principle more tractable but are not available in the present system. Here, we compare the simulation with experiment on the basis of secondary structure content. Fig. 4 *b* shows the evolution of the domain secondary structure content during 450K-U simulation, determined using STRIDE (Frishman and Argos, 1995). The extent and stability of the different helix segments are clearly in remarkable agreement with the experimental data. 1) The full B-helix, residue D₂₁ (N-cap) to S₃₁, remains virtually stable with the native content along the full trajectory and with only small fluctuations mainly at the C terminus.

2) Experimentally, the A-helix, which in the native state extends from T₂ to K₁₆, is more unstable, with only the A₈-L₁₈ segment exhibiting helix propensity. This is also seen in the simulation. The N-terminal part of the A-helix, G₁-F₆, is mainly unfolded. Although STRIDE indicates D₉ as the first residue in the helix, A₈C=O also participates in the helix hydrogen bond network.

Residues A₈-L₁₈ in fact partially adopt a π -helix: α -helix equilibrium for much of the first one-half of the 450K-U simulation. After 1.4 ns this segment becomes mostly α . Subsequently, the C-terminal part of the A-helix is unstable, in agreement with the chemical shift index being lower than for the B-helix. These fluctuations also involve the nonnative helix prolongation K₁₆-L₁₈, which is described later.

Experimentally, the A-helix begins at D₇ with a nonnative capping box with E₁₀. In the simulation residue E₁₀ is sequestered by R₃₂ in a nonnative salt-bridge that replaces the native D₇-R₃₂ salt-bridge. Either the simulation is not long enough to allow the salt-bridge to break and the capping box to form, or the capping box is not stable at the present simulation temperature (see discussion below of pretransitional effects in the 300- and 350-K simulations).

The A-helix has also an interesting feature in the simulation at the C terminus: the helix is longer than the native helix by two residues, G₁₇-L₁₈ which form a particular C-terminal motif with an additional A₁₄-G₁₉ H-bond. This increase in helix length was experimentally observed as a $i - (i - 4)$ NOE contact from L18 side-chain in the AB fragment (Macquaire et al., 1993). This structure leads to suppression of the AB loop. This may be an example in which part of a protein is not at its local energy minimum (i.e., that it would adopt when isolated in solution) when in a complete domain. Heating the complete domain reduces some of the nonlocal constraints allowing the locally constrained regions of the protein to relax to their intrinsic energy minima.

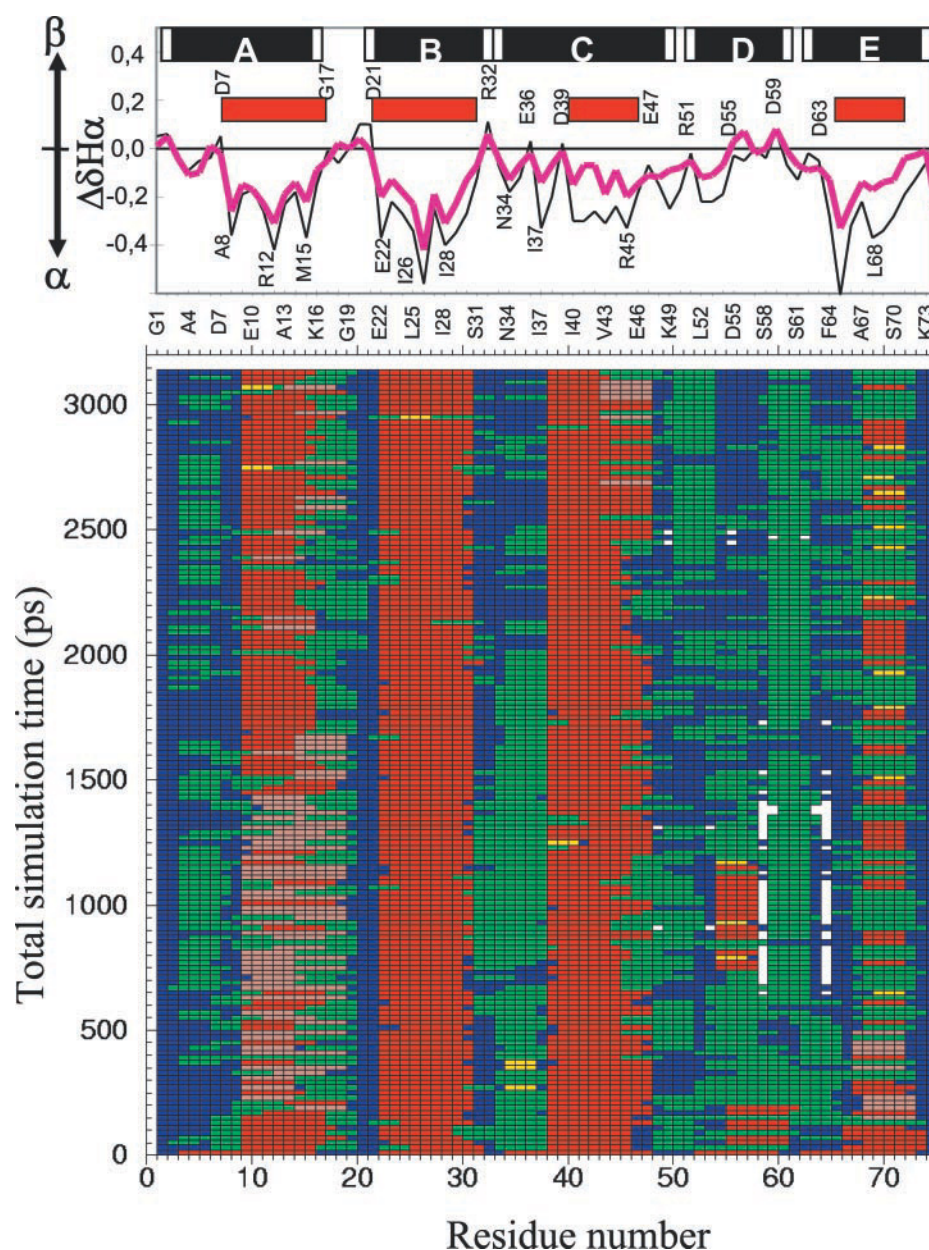


FIGURE 4 (a) NMR chemical shift data; (magenta heavy line) aqueous solution data; (black thin line) micellar solution data. The native helices are indicated as black boxes, and the residual helices in the unfolded state as red boxes; (b) Time series of STRIDE secondary structure: (red) α -helix; (yellow) 3_{10} helix; (magenta) π -helix; (green) turn; (blue) coil; (white) β -strand.

3) The E-helix, which in the native state extends from G₆₂ to K₇₃, also shows a transient π -helix structure at the beginning of the simulation, which is rapidly reduced to a little more than one helix turn in equilibrium with turn-like structures. The populations of these two states are approximately equal in accord with the experimental chemical shift indices in the E-helix region in pure aqueous solution being lower than for the other helices.

4) In the simulation the C-helix, which in the native-state extends from N₃₄ to L₄₈, is reduced to approximately the segment R₃₈-L₄₇. This is also true in the experiment. The C-terminal part, Y₄₄-L₄₇, of the residual C-helix exhibits large fluctuations in the simulation, thus reducing the stable part of the helix to approximately one-and-one-half helix

turns, again in agreement with the NMR data. The NMR data indicate that the residual helix structure is most probably N-capped by D₃₉. This correlates with high stability of the residual helix N terminus in the simulation. However, in the simulation D₃₉ does not actually cap this residual helix but rather makes a quasipermanent intraresidue side-chain-to-backbone H-bond. This kind of interaction is frequently observed in the simulation, other examples being D₇ in the A-helix and D₂₁ at the N terminus of B-helix, and produces a local constraint on the backbone conformation similar to that provided by capping.

The N terminus of the C-helix, N₃₄-D₃₉, was shown experimentally to adopt persistent nonnative (probably turn-like) structures. This was clearly demonstrated by oscilla-

tions in the chemical shift index amplified by the micellar effect. In the present simulation, this segment presents no helical structure. The MD N₃₂-D₃₉ segment is a slowly fluctuating loop in which the three residues R₃₂, E₃₆, and D₃₉ have approximately constant relative distances. The experimentally observed nonnative R₃₂-D₃₉ salt-bridge is not observed in this simulation (450K-U) but is observed at the last 600 ps of the 450K-B simulation. Thus, the present 3-ns 450K-U simulation, even at 450 K, is likely to be too short to be able to reproduce this conformational exchange.

5) In the simulation the segment E₄₇-A₆₇, spanning the C-helix C terminus, the D-helix, and the E-helix N terminus, is partially disordered and even transiently shows a small β sheet. Turn structures are apparently well populated. This is also clearly reflected in the ¹H chemical shift index profile. The positive or null values observed in the D-helix indicate a population of extended structure and no helix. This feature is also well reproduced in the simulation: the segments K₅₄-T₅₇ and S₅₈-T₆₀ remain in a quasi-extended conformation during the two last nanoseconds.

Fig. 3 *a* shows the evolution of the 450K-U structure after the 145-ps helix repulsion constraint phase. The initial global topology is extended with a high radius of gyration. The structure at the end of the 450K-U simulation is depicted in Fig. 3 *b* in which three snapshots over the final 500 ps are shown with the A- and B-helices superposed. Remarkably, the domain recovers the essentials of the native topology, i.e., the winding of the five segments corresponding to the five native helices into an approximate right-handed super-helix. This is the result of the existence of a stable structure, composed essentially of the AB helix hairpin substructure (residues D₇-R₃₂) about which the E-helix moves. At the same time the C-D segment fluctuates as a single flexible segment. The E helix continually makes hydrophobic contacts with A-B (mainly with B). These results are also consistent with the NMR ¹H chemical shift index profiles of the whole domain and of the ABC and DE fragments, which showed that the helix population in the E segment is significantly higher in the complete domain than in the DE fragment, presumably as a result of the helix-helix interactions DE fragment (Cordier-Ochsenbein et al., 1998a).

The ABE structure may well form a folding framework for the annexin 2 domain. The framework is composed of the domain segments having the highest intrinsic helix propensity. The AB helix hairpin, i.e., segment D₇-R₃₂, constitutes the most stable part of the protein. This is consistent with the experimental fragment studies, which showed that the AB fragment substantially populates a helix hairpin structure, in contrast to the other fragments e.g., CDE, which are much less structured (Cordier-Ochsenbein et al., 1998a). The E-helix may maintain its structure by persistently interacting with this framework. The compactness of the stable ABE framework is evidenced by its radius of gyration, which remains close to the native-state value, ~10 Å in most of the simulation frames.

In conclusion, the present simulation protocol reproduces the experimental NMR data on the equilibrium-unfolded state of the annexin I domain 2 with remarkable accuracy, including many details. This provides strong support for the conclusions drawn from the analysis of the unfolding trajectories, which are described in the following.

Exploring the native state basin, N

This region is explored by four independent simulations: for ~2 ns at 300, 350, and 400 K and during approximately the first 3 ns of the common part of 450K-A and 450K-B. In the 450-K simulation, the native state basin exploration clearly represents a lag-time before unfolding that is significantly longer than in hitherto published unfolding trajectories, which have often been unrealistically nearly “explosive,” occurring within ~0.1 ns.

The native region is characterized by relatively small fluctuations of the RMSD from the crystal structure (which remains under 3 Å), the radius of gyration, R_g , and the accessible surface area (ASA) (Figs. 2, 5, and 6). In contrast, during this period there are considerable fluctuations in the fraction of native contacts Q and Q_c , the fraction of native contacts of the residues constituting the native protein core (A₈, L₁₁, M₁₅, L₂₅, L₂₉, I₃₇, I₄₀, Y₄₄, L₅₂, I₅₆, L₆₈, and A₇₂). This fluctuation period begins with the increase in the core residue RMSD and radius of gyration, R_{gc} shown in Figs. 2 and 6 at 1 ns and is indicated in cyan in Figs. 2, 5, 6, and 7.

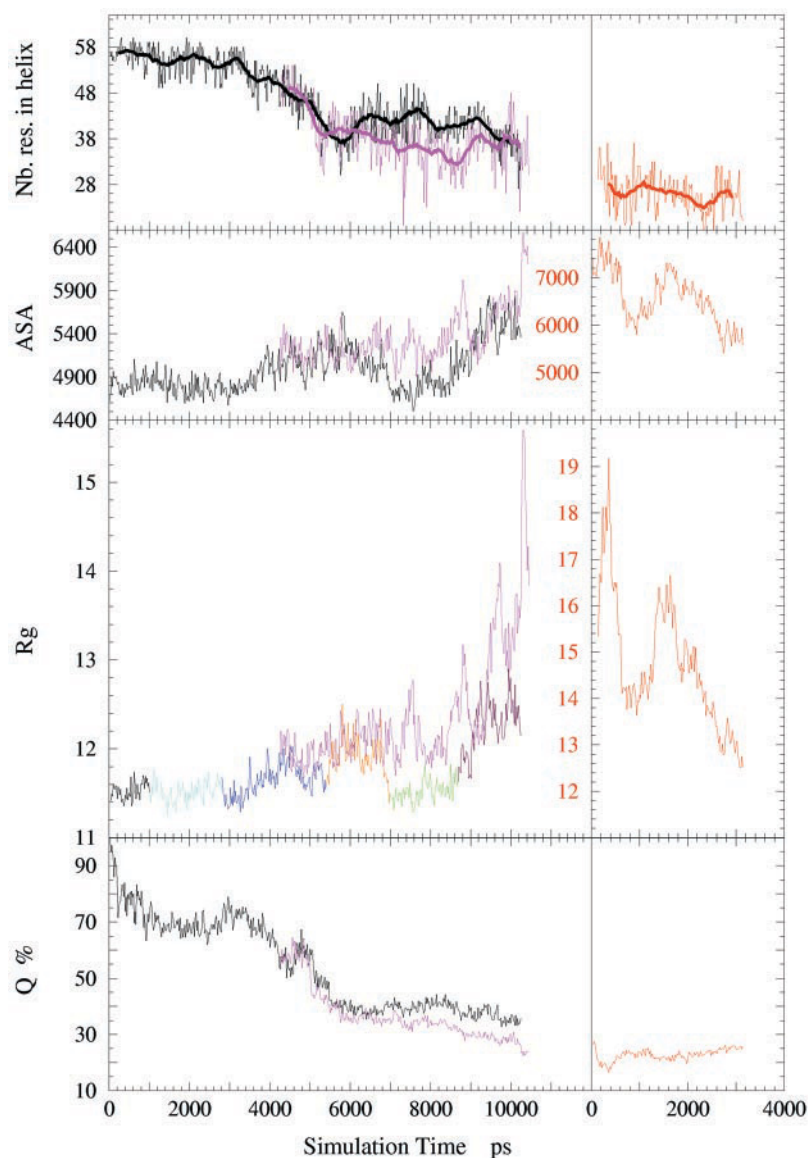
After 1.8 ns the protein has lost ~30% of its native contacts and the core 20% (30 of 132). Remarkably, Q_c returns to its native value (100%) at ~2.8 ns. Thus, the core refolds to the native structure. However, close to 30% of the noncore native contacts remains lost. The core refolding is therefore accompanied by irreversed structural changes in the peripheral residues.

The fluctuations of the protein up to 2.8 ns are nonproductive for core unfolding and are therefore considered as “pre-translational,” i.e., preceding the transition to the intermediate compact state. The pretranslational fluctuations exhibit both local and extended behavior. The local fluctuations appear as conformational equilibria. The sequence-dependence of the RMS fluctuation in 2-ns simulations performed at four temperatures (300–450 K) are given in Fig. 8. The pretranslational fluctuations occur principally in three regions of the protein as follows.

N-terminal part of the A-helix

The residues T₂PAQF₆ alternate between a single turn of an α -helix structure (T₂-F₆) and a type I turn (T₂-Q₅) at 300 K. Fig. 9 *a* shows an important 1 to 4 helix H-bond distance. At 300 K the helix is mostly stable with a few transient breaks in the 1 to 4 H-bond, which then forms a 1 to 3 H-bond characteristic of a type I turn. At 350 K the helix unfolds at

FIGURE 5 Number of residues in helix, accessible surface area ASA, total gyration radius R_g , and total native contact Q time series for the three trajectories: (black, cyan, blue, orange, green and brown) 450K-A trajectory; (magenta) 450K-B trajectory; (red) 450K-U trajectory. For the R_g data the 450K-A trajectory is separated into six time periods indicated by different colors: (cyan) fluctuation period in the native basin; (blue) transition period T1; (orange) strong fluctuation period preceding the formation of the kinetic trap structure; (green) kinetic trap period; (maroon) end of the 450K-A trajectory.



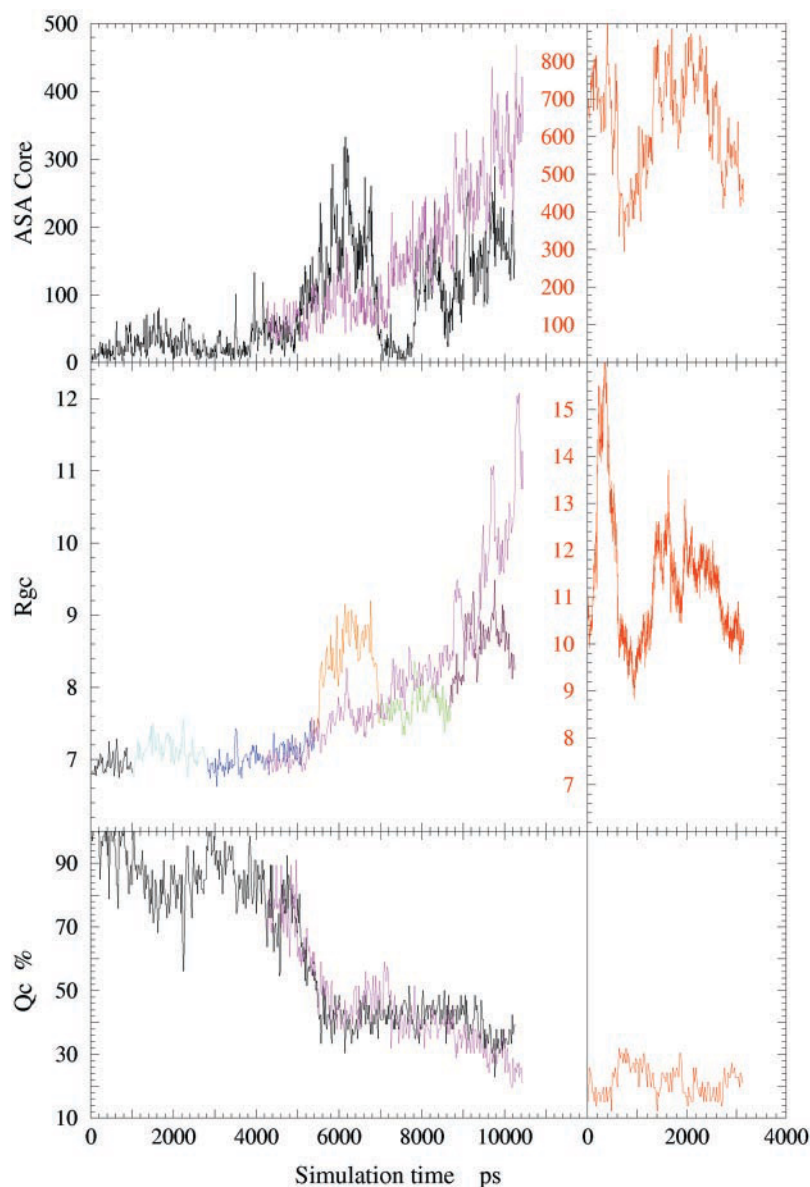
550 ps and refolds again at ~ 1 ns, reforming the native capping box. At higher temperatures the helix loses structure irreversibly.

A₁₄-L₁₈ segment and the N-terminal segment of the B helix

The AB loop is incorporated into the C terminus of the A-helix. This involves obligatory inversion of the K₁₆-G₁₇ peptide plane involving crossing of a local conformational barrier and associated formation of an A₁₄-L₁₈ H-bond. Fig. 9 *b* shows the A₁₄-L₁₈ H-bond distance time series at the explored temperatures. The conformational energy barrier is high enough at the lower temperatures to block the transition during the ~ 2.5 -ns simulation time. However, the transition occurs early (~ 200 ps) in the 450 K trajectory, and the H-bond remains relatively stable thereafter.

The AB loop fluctuations are associated with those of the N-terminal part of the B-helix. This helix exhibits high stability both in fragment studies with NMR (Macquaire et al., 1992) and in the present unfolded-state simulation (450K-U, see later description). However, in the pretransitional phase it is subject to large-amplitude fluctuations as attested by numerous hydrogen-bond breaking events in the residues preceding I₂₈ shown in Fig. 9 *c*. The reason for these large fluctuations may lie in the fact that in the intact protein the B helix is strongly involved in the interdomain interface (see Fig. 1), which was removed at the beginning of the present nonequilibrium unfolding simulations. In contrast, the D-helix (which is the least stable helix in the NMR fragment experiments and in the unfolded state 450K-U simulation) and the C-helix do not present any substantial fluctuations during the pretransitional phase.

FIGURE 6 Core residue ASA, core residues radius of gyration R_{gc} , and native contact for core residues Q_c time series. The colors are the same as for Fig. 5.



E-helix

As well as typical fraying of the termini, this helix presents a rigid-body motion, moving around the D-E pair. This motion leads to the increased fluctuations in Fig. 6 and the variations in the L_{52} - A_{72} $C\alpha$ distance in Fig. 9 *d*.

The above ensemble of well-identifiable conformational fluctuations constitutes an exploration of the conformational space that can be thought of as a search for a transition pathway to a intermediate compact state, which would involve new fluctuations capable of cooperatively driving the protein to the transition region.

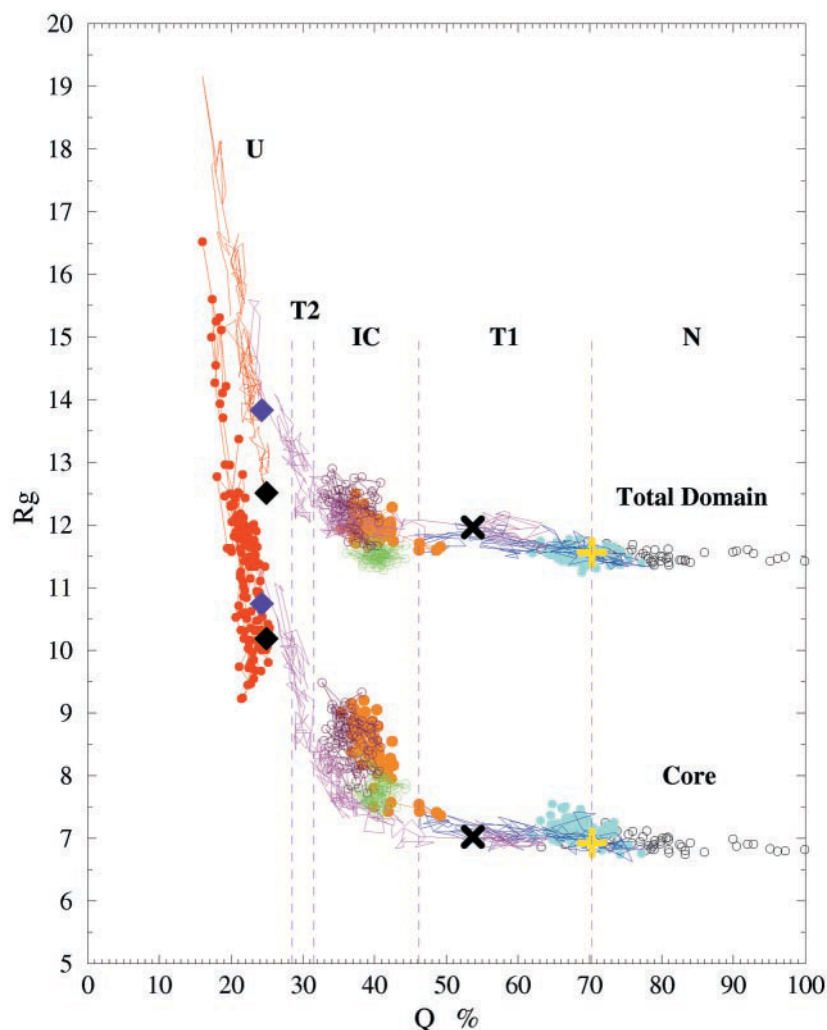
Transition region T1

Figs. 5 and 6 indicate that this transition region lasts ~ 2 ns (from ~ 3.5 to ~ 5.5 ns) and involves a large reduction in Q

and Q_c . Despite the large variation of Q the increase in the radius of gyration relative to the native state is rather modest, at ~ 0.5 Å. The bifurcation point of 450K-A and 450K-B is located in this transition region.

It is of particular interest to find key moments along the unfolding trajectory in which transient events lead to significant unreversed structural change. At the beginning of the simulation, during the first 3 ns, the pretransitional fluctuations appear to be independent from each other (see Figs. 2 and 9). Subsequently unreversed conformational changes occur, engaging the protein on an unfolding track. The RMSD data of Fig. 2 show that the first large fluctuation, which is not reversed before another part of the protein unfolds, concerns the B-helix at ~ 3.5 ns. Fig. 10 shows time series of the backbone angles of K_{16} and G_{17} . Clearly, the 3.5-ns helix unfolding accompanies the nearly simulta-

FIGURE 7 Total and core radius of gyration, R_g and R_{gc} versus Q . Same colors as Fig. 5 and 6. (Yellow +) B-helix unfolding point (see text); (black \times) bifurcation point for 450K-B trajectory; (violet diamond) end of the 450K-B trajectory; (black diamond) end of the 450K-U trajectory.



neous inversion of the two peptide planes K_{16} - G_{17} and D_{23} - T_{24} . 450K-A and 450K-B are closely related, even one nanosecond after the bifurcation. This point is indicated as a yellow “+” in Fig. 7. The end of the T1 transition period for both trajectories occurs at ~ 5.4 ns (for example, see the Q_c /time plot of Fig. 6). At this time the C, D, and E-helices have partially or completely unfolded (see RMSD of these helices on Fig. 2).

Compact partially unfolded or intermediate compact region, IC

In both 450K-A and 450K-B the end of the transition leads to a intermediate compact state, which is associated with the almost complete unfolding of the D helix. The intermediate compact region (~ 5.4 – 8 ns) involves strong fluctuations of the radii of gyration (Figs. 5 and 6) and a corresponding expansion of the protein of 13% for R_g and 28% for R_{gc} .

The two 450-K trajectories behave quite differently. In 450K-A the protein remains in the intermediate compact

state until the end of the simulation. In 450K-B the protein remains in the intermediate state for ~ 3 ns and then enters an unfolded state, accompanied by additional partial unfolding of the C-helix. Moreover, the 450K-B trajectory exhibits a gradual increase in the global parameters R_g and ASA, whereas 450K-A shows somewhat more complex behavior (see the three colored time periods in Figs. 5 and 6). Consequently, although both trajectories 450K-A and 450K-B cluster in the same region of the R_g/Q plot (labeled CI) they occupy distinct subclusters (Fig. 7): the R_{gc} values for 450K-B are on average associated with smaller Q values than those of 450K-A.

In 450K-A three time periods can be distinguished, which are given three different colors in the figures (orange, green, and maroon, successively). 1) 450K-A enters the intermediate compact state with a large (~ 2 Å) increase in the core R_g . During this first period (~ 5.4 ns to ~ 6.9 ns, orange in Figs. 5–7) there is also compaction of the ABC subdomain into a three-helix bundle accompanied by slow diffusion of the E-helix N terminus over the ABC surface (Fig. 11 a).

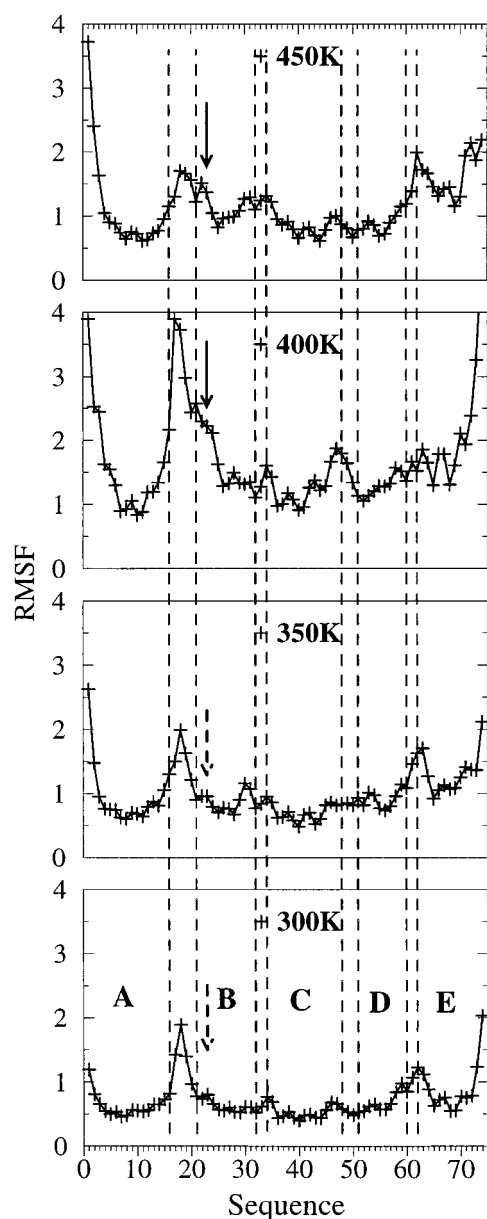


FIGURE 8 RMS fluctuation of C α atoms versus sequence at four simulation temperatures in the native basin. Fluctuations affecting the T20-T24 segment are indicated by an arrow.

During this diffusion the E-helix strongly fluctuates as an approximately rigid body, transiently exposing its core residues to the solvent. These transient exposures dominate the large-amplitude fluctuations of the core ASA, the orange period visible in Fig. 6. 2) The first period ends with a relatively fast (≈ 100 ps) return of the E-helix into contact with the ABC subdomain. This forms a nonnative tertiary structure stable between 7 and 8 ns (green in Figs. 5–7) for which, remarkably, R_g is as small as that of the native state. However, R_{gc} remains larger than the native value indicating that the core residues have been rearranged. This tran-

sient “very compact” intermediate structure comprises the ABC, the topology of which has become a three-helix bundle shown in Fig. 11 *a*, a single helix turn L₅₂–L₅₆ replacing the D-helix, and the E-helix in the α conformation oriented perpendicularly to the bundle. The new super-helix topology in the intermediate is more highly wound than in the native state, with a $\approx 630^\circ$ total rotation of the helical segments, compared with $\approx 540^\circ$ for the native state. Fig. 11 *b* shows the structure of the very compact intermediate core residues. The number of core residues is smaller than that in the native core, being composed of only eight residues (L₁₁, M₁₅, L₂₅, L₂₉, V₄₀, Y₄₄, L₆₈, and A₇₂) instead of 12 residues. This reduced core contains residues from helices A, B, and E plus two residues of the small remainder of the C-helix. The reduced core constitutes, in the intermediate compact state, a stable subset of residues. However, cavities are observed in the core (Fig. 11 *b*), which is thus less densely packed than the native state core.

We call R_{gci} the radius of gyration of the reduced core residues in the intermediate state. The R_{gc} and R_{gci} time series are given in Fig. 12. R_{gc} and R_{gci} are closely similar in the native state, the transition ensemble and unfolded state at the end of 450K-B but differ considerably in the intermediate compact ensemble. In the intermediate compact ensemble R_{gci} remains small and essentially equal to the native value (≈ 7 Å) for both runs 450K-A and 450K-B. This shows that, although there is no transiently stable very intermediate compact in 450K-B, the residues corresponding to the 450K-A intermediate core (7.0–7.7 ns) have the same radius of gyration in 450K-B. This reveals the existence in the intermediate compact state of a subset of residues of closely similar compactness in both trajectories.

The structure of the very compact intermediate in 450K-A is stabilized by a large number of salt bridges and side-chain-to-backbone H-bonds, also shown in Fig. 11 *b*. Sixteen residues make 10 interaction pairs including nine salt bridges, contrasting with only two salt-bridges observed in the native state. Only one of these nine salt-bridges is native, the conserved D₇–R₃₂ interaction. The nonnative salt-bridge network links the A and C helices and involves five residues: two arginines, two glutamates, and one aspartate. The electrostatic interactions are between surface residues, possibly contributing to the compactness of this structure. 3) Between 9 ns and the end of 450K-A (10.1 ns) R_g is increased and more highly fluctuating.

Hydration in the transition region T1 and the intermediate compact state

The hydration of the core residues, as represented by the number of water molecules within 2.5 Å with any core residue atom, is given for the three 450-K trajectories in Fig. 13. The native state core is in contact with 4 ± 2 water molecules involving accessibility of three of the core residues. In the transition region T1 there are on average only

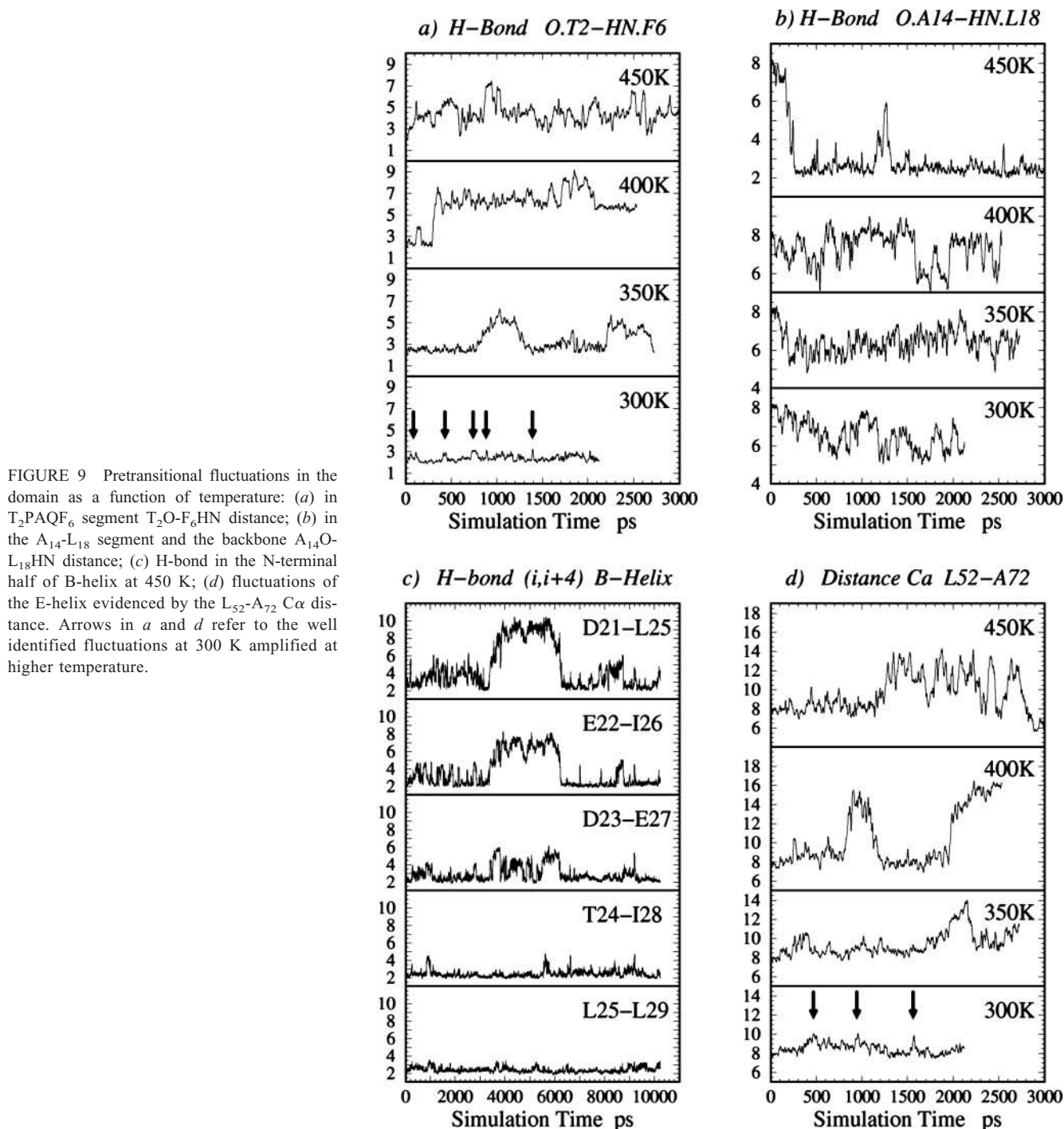
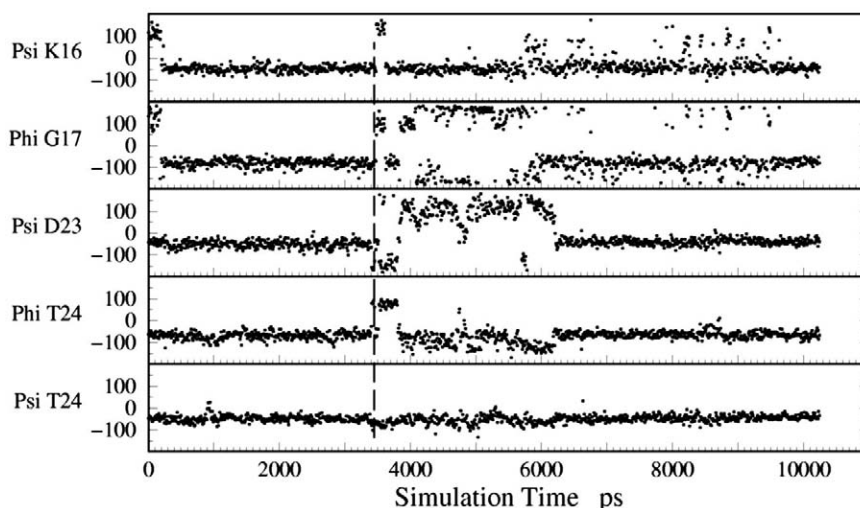


FIGURE 9 Pretransitional fluctuations in the domain as a function of temperature: (a) in T₂PAQF₆ segment T₂O-F₆HN distance; (b) in the A₁₄-L₁₈ segment and the backbone A₁₄O-L₁₈HN distance; (c) H-bond in the N-terminal half of B-helix at 450 K; (d) fluctuations of the E-helix evidenced by the L₅₂-A₇₂ C α distance. Arrows in a and d refer to the well identified fluctuations at 300 K amplified at higher temperature.

two additional water molecules in contact with core residues. This small increase in the core hydration is accompanied by a 6% increase in the total ASA, which is associated with the unfolding of the K₁₆-T₂₄ segment and various helix termini. This surface unfolding itself increases the exposure of surface residues, increasing the number of water molecules in contact with the protein surface but not with the core. The large fluctuations in 450K-A (5–7 ns) of the core

ASA and hydration number arise essentially from rigid-body fluctuations of the E-helix during its diffusion on the ABC core surface (Fig. 8). This involves solvent exposure of the E-helix, facilitating the formation of the ABC three-helix bundle. Some hydrophobic side-chains that are exposed or partially exposed in the native state become more buried in the intermediate compact structure, thus compensating for the exposure of the E-helix initial core residues.

FIGURE 10 $\phi\psi$ time series for G17, D23. The dashed line indicate the correlated transition at 3.5 ns corresponding to the yellow + in Fig. 7.



Closer inspection of the core fluctuations in the intermediate compact state revealed a number of transiently formed, “flickering” cavities that increase the core ASA (Fig. 6, 5–7 ns). Examples of these cavities are given in Fig. 14. Interestingly, the cavities were never hydrated. The cavity life-

times may not be long enough to permit water to diffuse in. Furthermore, the cavities were found to open mostly near distal polar side-chains, which presented a further barrier, by hydrogen bonding to close water molecules. In some cases, water molecules were found to be transported by polar side-chains to which they are H-bonded (especially glutamine), back and forth from the surface to within contact of the core residues.

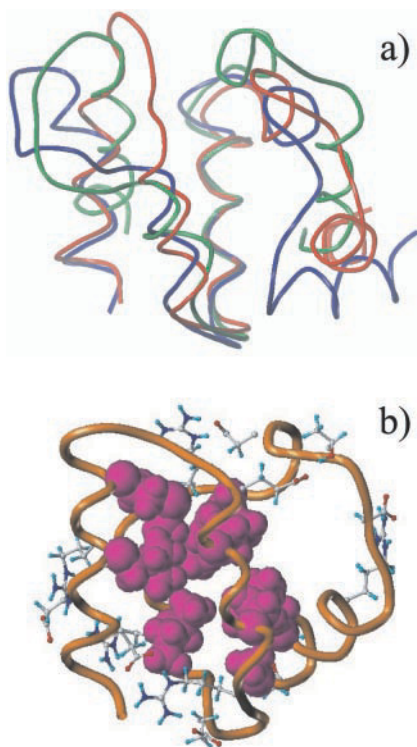


FIGURE 11 (a) Superposed structures showing diffusion of E-helix on ABC three-helix bundle in the 450K-A trajectory that finishes by the formation of the conformational trap; (green) 5 ns; (blue) 5.9 ns; (red) 7 ns; (b) Very compact intermediate (trap), found between 7 and 8 ns in 450K-A. The hydrophobic core residues are shown as space filling atoms, and the extensive salt-bridge network is also indicated. Cavities in the core are visible.

Transition region, T2

In 450K-A the domain does not unfold and remains in the intermediate compact state up to the end of the simulation. In 450K-B, however, further unfolding is seen. Fig. 12 allows an estimation of the lifetime of the intermediate compact state in 450K-B. R_{gc} and R_{gci} first diverge at the end of T1 then converge at ~ 9.4 ns when, after large fluctuation, R_{gci} exhibits a fast increase. This behavior allows T2, the intermediate compact to unfolded state transition, to be detected. Using data of Figs. 2, 5, and 13 this second transition can be tentatively identified as taking place between ~ 8.7 and 9.5 ns. Interestingly, then, whereas the T1 transition is characterized by a large loss of native contacts, a very small expansion and a negligible hydration of core residues, T2, in contrast, exhibits only small decreases of Q and Q_c but a large expansion and increased hydration.

DISCUSSION

Our earlier experiments showed that the equilibrium state of domain 2 combines a large folding framework with native or quasative interactions, nonnative local structures, and completely unfolded segments. The present simulation now provides a more precise picture of this unfolded state. Although expanded, the state is more compact than would be

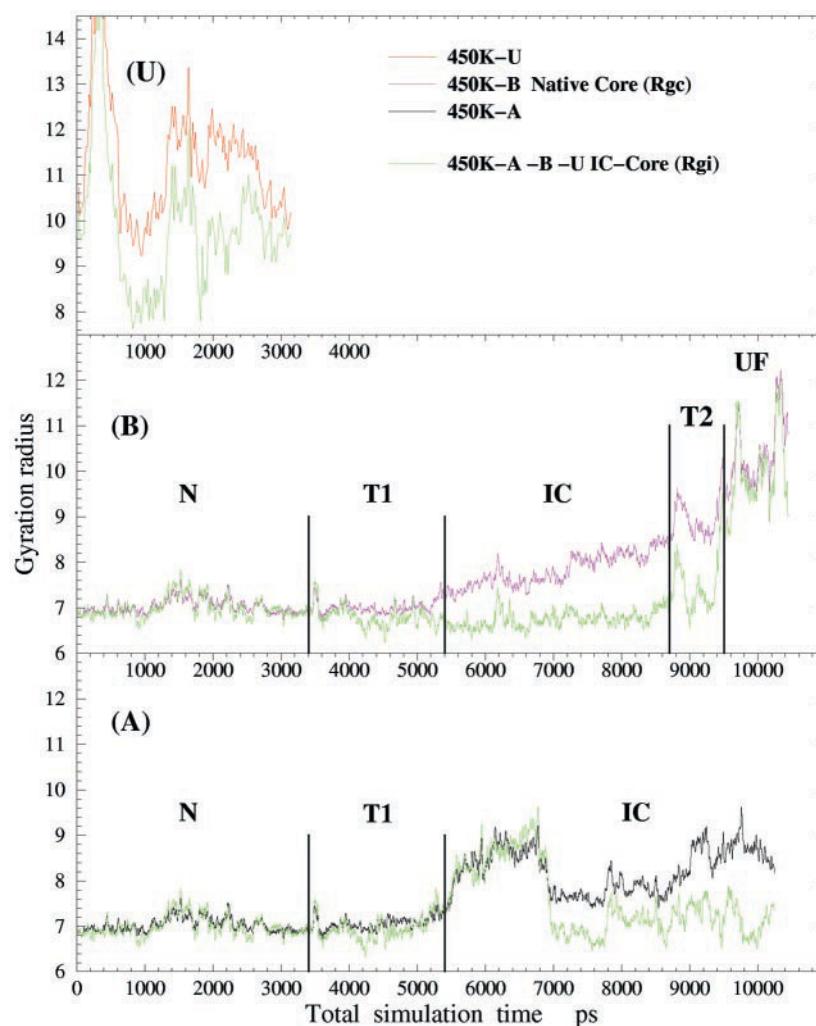


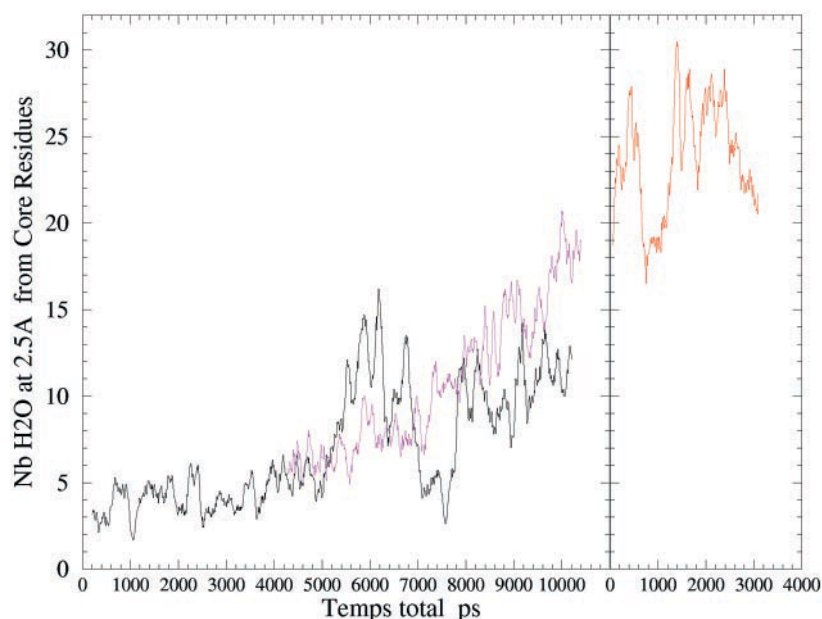
FIGURE 12 Native core radius of gyration R_{gc} (black, magenta, and red) and intermediate reduced core radius of gyration R_{gci} (green) time series for the A, B, and U simulations.

expected for a fully random chain (Petrescu et al., 1998, 2000). It is also less expanded than several other small proteins studied with simulation (Wong et al., 2000; Mayor et al., 2000; Prompers et al., 2001). Moreover, the simulation suggests a mechanism by which a residual native-like topology can be maintained in the disordered structure: the hydrophobic face of the AB hairpin would offer the necessary surface on which the E-helix can loosely interact. The E-helix diffuses over the AB hydrophobic face, increasing the protein chain entropy relative to a more rigid folding framework. This may be a means of finding a compromise between the necessity of having a folding framework, with a reduced average radius of gyration, and the chain entropy cost on folding.

As the domain 2 is not intrinsically stable, it is necessary to place it in the context of the unfolded entire four-domain protein when considering the reverse reaction, i.e., domain 2 refolding. As is the case with domain 2, domain 3 is not an intrinsically stable domain, whereas in contrast domain 1 is intrinsically stable. Our earlier hypothesis (Cordier-Ochsenbein et al., 1998b) was that this instability is related

to the necessity of keeping a sufficiently high level of flexibility in both domain 2 and 3 so as to facilitate the conformational search for docking domain 4 to the first-folding domain, domain 1. In this model, folding of domain 2 is triggered by domain 2-domain 4 polar interactions (salt-bridge locking) as well as by interaction with domain 3. Folding of both domains is consolidated when they are able to establish the four-helix bundle that makes their hydrophobic interface. Interestingly, the folding framework of domain 2 as seen both by NMR and in the present simulation consists of the three helices that form both interfaces: AB with BA of domain 4 and BE with EB of domain 3 (see Fig. 1). Although the model needs further independent experimental testing, for example, by protein engineering, the present simulations do provide additional support for it by showing that a native-like topology can be maintained in the unfolded isolated domain 2 despite the loss of secondary structure. Moreover, the simulations also show that when the topology of the domain is strongly distorted it can, after large amplitude fluctuation, relax back to a native-like topology (Fig. 3). As already mentioned, the

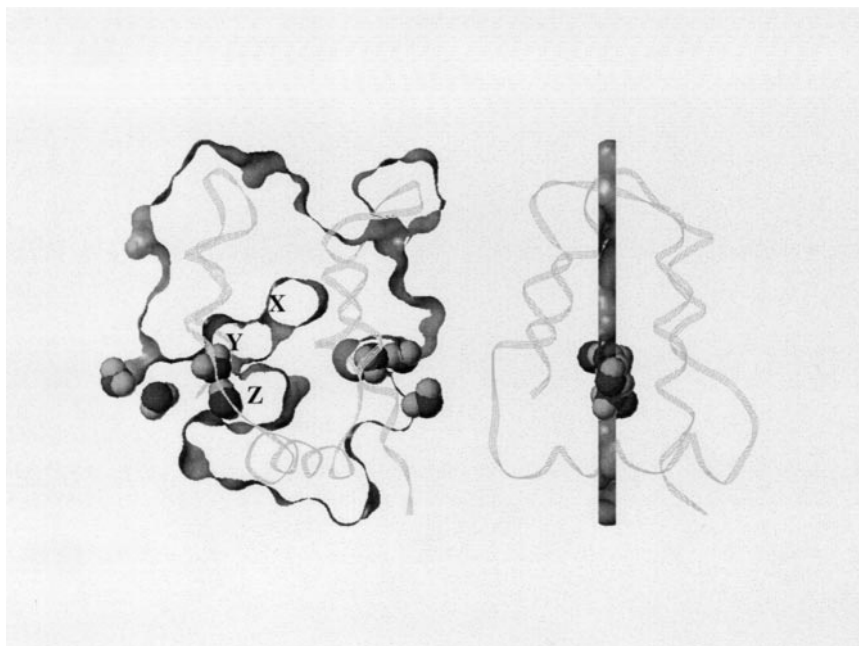
FIGURE 13 Hydration of domain given by the number of water molecules within 2.5 Å of the core residues for the three trajectories (same color as in Fig. 2).



near-equilibrium unfolded state is more compact than the unfolded state of some small proteins. This is important because once the domain 1 is folded, the compactness and the initially correct topology in domain 2 will not only accelerate its own folding but also will help the protein to acquire rapidly the correct circular organization of the remaining domains. Recent work (Chiti et al., 1999; Martinez and Serrano, 1999; Riddle et al., 1999; Ternstrom et al., 1999) has emphasized the major role of native topology in folding reactions.

It remains to be understood why the domain 2 sequence is intrinsically not foldable whereas domain 1 is. We have suggested that this difference may arise from the presence of various salt-bridge-stabilized nonnative structures experimentally observed in isolated domain 2 that could have the effect of stabilizing the unfolded with respect to the folded state (Cordier-Ochsenbein et al., 1998a). Because the exchange frequency between these different nonnative structures is likely to be rather low (e.g., microsecond to millisecond time scale) the simulation is obviously unable to

FIGURE 14 Example of nonhydrated cavities in T1(X, Y, Z): two orthogonal representations of a Connolly surface slice of the domain ($R_{\text{probe}} = 1.4$ Å). The water molecules at 2.5 Å from domain atoms present within the slice are shown with the space filling representation.



confirm this point, although some of the nonnative salt-bridges are indeed observed. Nevertheless, the simulation draws our attention to an additional possibility. Returning to the observation that the pretransitional fluctuations involve essentially the folding framework, even at 300 K, we may suggest that the packing of the ABE helices and perhaps also their packing on the C and D helices is simply not sufficiently strong to ensure the stability of the isolated native structure and needs the additional stabilizing interactions with the other domains, particularly domain 3. The packing of the E-helix seems to be important in the annexin domain topology and E-helix rigid-body fluctuations were also observed in the 300 K simulation of the domain 1 (Huynh et al., 1999). However, in the case of the domain 1 simulation the E-helix fluctuations led to a stable repositioning of this helix with a different packing of its core residues on the rest of the protein. This repositioning, which also lead to a reduction of the hydrophobic ASA, may partly explain the stability of this domain.

The present results reveal a rich variety of conformational behavior during a high temperature nonequilibrium molecular dynamics simulation of the unfolding an intrinsically unstable protein domain. The roughness of the native energy landscape is such that the protein has to cross a large number of energy barriers separating the different minima before being able to productively unfold: the domain exhibits considerable local and global pretransitional fluctuations before finding an effective unfolding pathway. These fluctuations are not evenly distributed over the whole protein. Thus, it may be that there exists an obligatory pathway from the native state to transition region during the initial steps of the thermal unfolding of the domain. The pretransitional fluctuations may have the consequence of accumulating a critical number of small structural defects in the protein core, transiently stabilizing nonnative interactions, and allowing the protein to enter into the T1 transition region.

The pretransitional fluctuations concern the folding framework (ABE helices) rather than the helices with the lowest intrinsic helix propensity (helices C and D). The folding framework is composed of precisely the helices that interact with the domain 3 (B and E) and with the domain 4 (A and B) in the entire protein. Therefore, the increased ABE fluctuation in the present nonequilibrium simulation may be due to the fact that the domain interface with ABE has been removed.

In 450K-A a particularly interesting transient very compact state is found at 7.0 to 8.8 ns. In 450K-B there is no evidence of such a state, and the protein unfolds more than in 450K-A. The 450K-B simulation might therefore correspond to a downhill unfolding of the domain, whereas the very compact state in 450K-A is not an obligatory intermediate but rather an example of a kinetic trap. Consequently, in 450K-A the transition between the intermediate compact state and the unfolded state is not observed. For this to happen the simulation would have had to be long enough for

the three-helix bundle to unfold. The existence of fast and slow tracks in folding/unfolding reactions lends clear support to the energy landscape description. Folding trajectories slowed by kinetic traps have been observed in lattice simulations in parallel with fast folding trajectories (Dill et al., 1995). However, to our knowledge the present result is the first time that a structurally well-defined unfolding trap has been observed in an all-atom simulation.

The R_g expansion in the intermediate compact state ($\sim 13\%$) falls within the lower end of the range of values observed by small-angle x-ray scattering for the molten globule states of several proteins (Kataoka et al., 1997; Segel et al., 1998). The range of the fraction of native contacts, $45\% > Q > 30\%$, also corresponds to the molten globule values observed in lattice simulations (Sali et al., 1994; Shakhnovich, 1998). For these reasons the intermediate compact region observed here may be tentatively assigned as a molten-globule state of the annexin domain.

The structure in the intermediate compact state still contains 60% to 70% of native secondary structure. The loss of secondary structure in this state now affects essentially the segments with experimentally determined low helix propensity, namely the C and D helices, which continuously unfold to a degree eventually close to that found in the unfolded state. Interestingly, the B-helix, which is partially unfolded at the beginning of the transition zone T1, refolds during the intermediate compact state-to-unfolded state transition, T2. This suggests that interactions working against intrinsic secondary structural propensities are still present in the intermediate compact state and are relaxed only in the unfolded state.

Identifying the intermediate compact state as a molten globule we recall that this state has been thought of as relatively highly hydrated. However, recent considerations and NMR experiments on different molten globules suggested and demonstrated that the molten globule state is in fact poorly hydrated (Nath et al., 1996; Denisov et al., 1999). The presently simulated molten globule is clearly devoid of significantly increased hydration. The water molecules do not quickly diffuse into the structure so as to cause unfolding by dissociation of the core residues. In contrast, hydration of the core residues is a more or less continuous process during the entire simulation: native core residues become exposed to the solvent because of the progressive plastic deformation of the structure.

It will be interesting to see if low compact-intermediate hydration is a common property of purely helical proteins having at least one high propensity amphipathic helical segment as a folding framework. In these cases the amphipathic helices may form an early hydrophobic surface favoring early hydrophobic side-chain collapse, rapidly stabilizing the intermediate compact state. In the case of the annexin domain the folding framework is rather large as it is composed of approximately three-fifths of the domain. The presence of this large water-excluding folding frame-

work in the intermediate compact state may explain why the hydration increase of this state is very small compared with the native structure. In this context it would be interesting, using our simulation protocol, to study the unfolding transition of a small protein with a large amount of β -structure.

To end, it is worth observing the clustering of the two trajectories for the intermediate compact state as revealed by the R_g/Q plot of Fig. 7. This clustering is thought to reveal the presence of several local minima in the energy landscape of the protein. One question is how many other minima exist that are not explored because of the limited number of trajectories presently calculated. In our opinion, because the topology of the domain is strongly preserved, even in the unfolded state, these other local minima should be very close to those already observed. Their exploration by the unfolding protein should correspond to small fluctuations in the main unfolding pathway that precedes the second unfolding transition T2. Only after these second transition will the unfolding pathways more strongly diverge and will need much more extensive simulation to be well represented. This representation is still out of range of our present time simulation power.

In conclusion, then, the present work has revealed a variety of states and transitions of the annexin domain on the way to unfolding in nonequilibrium simulation at 450 K. Although the timescale explored (10^{-8} s) is several orders of magnitude smaller than that of protein folding at 300 K, this is somewhat compensated by the high temperature used which accelerates the barrier-crossing processes involved. Furthermore, although the duration of the simulation of the unfolded state (3 ns) even at 450 K is unlikely to be sufficient to completely explore the associated conformational space clearly, the present unfolded state simulation reproduces the experimental NMR data on the equilibrium-unfolded state with remarkable accuracy. This gives confidence in the reliability of the present methodology and in the pertinence of the brief but richly detailed glimpses of unfolded and unfolding proteins that can be provided by molecular dynamics simulation.

REFERENCES

- Boczko, E. M., and Brooks, C. L., 3rd. 1995. First-principles calculation of the folding free energy of a three-helix bundle protein. *Science*, 269: 393–396.
- Brooks, B. R., R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. 1983. CHARMM: a program for macromolecular energy, minimization and dynamics calculations. *J. Comp. Chem.* 4:187–217.
- Chatfield, D. C., A. Szabo, and B. R. Brooks 1998. Molecular dynamics of staphylococcal nuclease: comparison of simulation with 15N and 13C NMR relaxation data. *J. Am. Chem. Soc.* 120:5301–5311.
- Chiti, F., N. Taddei, P. M. White, M. Bucciantini, F. Magherini, M. Stefani, and C. M. Dobson. 1999. Mutational analysis of acylphosphatase suggests the importance of topology and contact order in protein folding. *Nat. Struct. Biol.* 6:1005–1009.
- Cordier-Ochsenbein, F., R. Guerois, F. Baleux, T. Huynh-Dinh, A. Chafotte, J. M. Neumann, and A. Sanson. 1996. Folding properties of an annexin I domain: A H-1-N-15 NMR and CD study. *Biochemistry*. 35:10347–10357.
- Cordier-Ochsenbein, F., R. Guerois, F. Baleux, T. Huynh-Dinh, P. N. Lirsac, F. Russo-Marie, J. M. Neumann, and A. Sanson. 1998a. Exploring the folding pathways of annexin I, a multidomain protein: I. Non-native structures stabilize the partially folded state of the isolated domain 2 of annexin I. *J. Mol. Biol.* 279:1163–1175.
- Cordier-Ochsenbein, F., R. Guerois, F. Russo-Marie, J. M. Neumann, and A. Sanson. 1998b. Exploring the folding pathways of annexin I, a multidomain protein: II. Hierarchy in domain folding propensities may govern the folding process. *J. Mol. Biol.* 279:1177–1185.
- Daura, X., W. F. van Gunsteren, and A. E. Mark. 1999. Folding-unfolding thermodynamics of a beta-heptapeptide from equilibrium simulations. *Proteins*. 34:269–280.
- Denisov, V. P., B. H. Jonsson, and B. Halle. 1999. Hydration of denatured and molten globule proteins. *Nat. Struct. Biol.* 6:253–260.
- Dill, K. A., S. Bromberg, K. Yue, K. M. Fiebig, D. P. Yee, P. D. Thomas, and H. S. Chan. 1995. Principles of protein folding: a perspective from simple exact models. *Protein Sci.* 4:561–602.
- Duan, Y., and P. A. Kollman. 1998. Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. *Science*. 282:740–744.
- Eaton, W. A. 1999. Searching for “downhill scenarios” in protein folding [comment]. *Proc. Natl. Acad. Sci. U. S. A.* 96:5897–5899.
- Favier-Perron, B., Lewit-Bentley, A., and F. Russo-Marie. 1996. The high-resolution crystal structure of human annexin iii shows subtle differences with annexin V. *Biochemistry*. 35:1740–1744.
- Ferrara, P., and A. Caflisch. 2000. Folding simulations of a three-stranded antiparallel β -sheet peptide. *Proc. Natl. Acad. Sci. U. S. A.* 97: 10780–10785.
- Finkelstein, A. V. 1997. Can protein unfolding simulate protein folding? *Protein Eng.* 10:843–845.
- Frishman, D., and P. Argos. 1995. Knowledge-based protein secondary structure assignment. *Proteins*. 23:566–579.
- Gilquin, B., C. Guilbert, and D. Perahia. 2000. Unfolding of hen egg lysozyme by molecular dynamics simulations at 300K: insight into the role of the interdomain interface. *Proteins*. 41:58–74.
- Guerois, R., F. Cordier-Ochsenbein, F. Baleux, T. Huynh-Dinh, J. M. Neumann, and A. Sanson. 1998. A conformational equilibrium in a protein fragment caused by two consecutive capping boxes: 1H-, 13C-NMR, and mutational analysis [In Process Citation]. *Protein Sci.* 7:1506–1515.
- Huber, R., Berendes, R., Burger, A., Schneider, M., Karshikov, A., Luecke, H., Romisch, J. and E. Pâques. 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.
- Huynh, T., G. Musat, J.-M. Neumann, J. Smith, and A. Sanson. 1999. Molecular dynamics simulations of the isolated domain 1 of annexin I. *Theor. Chem. Acc.* 101:82–86.
- Ibragimova, G. T., and R. C. Wade. 1999. Stability of the beta-sheet of the WW domain: a molecular dynamics simulation study. *Biophys. J.* 77: 2191–2198.
- Inuzuka, Y., and T. Lazaridis. 2000. On the unfolding of alpha-lytic protease and the role of the pro region. *Proteins*. 41:21–32.
- Jorgensen, W. L., J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. 1983. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79:926–935.
- Kataoka, M., K. Kuwajima, F. Tokunaga, and Y. Goto. 1997. Structural characterization of the molten globule of alpha-lactalbumin by solution X-ray scattering. *Protein Sci.* 6:422–430.
- Lamy, A., and J. C. Smith. 1996. Denaturation of truncated staphylococcal nuclease in molecular dynamics simulation at 300K. *J. Am. Chem. Soc.* 118:7326–7328.
- Lazaridis, T., and M. Karplus. 1997. “New view” of protein folding reconciled with the old through multiple unfolding simulations. *Science*. 278:1928–1931.

- MacKerel, A. D. J., D. Bashford, M. Bellott, R. L. J. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. I. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, and M. Karplus. 1998. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B.* 102:3586–3616.
- Macquaire, F., F. Baleux, T. Huynh-Dinh, J. M. Neumann, and A. Sanson. 1992. ¹H-NMR conformational study of a synthetic peptide derived from the consensus sequence of annexins. *Int. J. Pept. Prot. Res.* 39:117–122.
- Macquaire, F., F. Baleux, T. Huynh-Dinh, D. Rouge, J. M. Neumann, and A. Sanson. 1993. Proton NMR conformational study of an annexin I fragment: influence of a phospholipidic micellar environment. *Biochemistry.* 32:7244–7254.
- Martinez, J. C., and L. Serrano. 1999. The folding transition state between SH3 domains is conformationally restricted and evolutionarily conserved. *Nat. Struct. Biol.* 6:1010–1016.
- Mayor, U., C. M. Johnson, V. Daggett, and A. R. Fersht. 2000. Protein folding and unfolding in microseconds to nanoseconds by experiment and simulation. *Proc. Natl. Acad. Sci. U.S.A.* 97:13518–13522.
- Nath, U., V. R. Agashe, and J. B. Udgaonkar. 1996. Initial loss of secondary structure in the unfolding of barnase [letter]. *Nat. Struct. Biol.* 3:920–923.
- Petrescu, A., P. Calmettes, D. Durand, V. Receveur, and J. C. Smith. 2000. Change in backbone torsion angle distribution on protein folding. *Prot. Sci.* 9:1129–1136.
- Petrescu, A. J., P. Calmettes, V. Receveur, and J. C. Smith. 1998. Excluded volume in the configurational distribution of a strongly-denatured protein. *Prot. Sci.* 7:1396–1404.
- Prompers, J. J., C. Scheurer, and R. Brüschweiler. 2001. Characterization of NMR relaxation-active motions of a partially folded A-state analogue of ubiquitin. *J. Mol. Biol.* 305:1085–1097.
- Riddle, D. S., V. P. Grantcharova, J. V. Santiago, E. Alm, I. Ruczinski, and D. Baker. 1999. Experiment and theory highlight role of native state topology in SH3 folding. *Nat. Struct. Biol.* 6:1016–1024.
- Sali, A., E. Shakhnovich, and M. Karplus. 1994. Kinetics of protein folding: a lattice model study of the requirements for folding to the native state. *J. Mol. Biol.* 235:1614–1636.
- Segel, D. J., A. L. Fink, K. O. Hodgson, and S. Doniach. 1998. Protein denaturation: a small-angle X-ray scattering study of the ensemble of unfolded states of cytochrome c. *Biochemistry.* 37:12443–12451.
- Shakhnovich, E. I. 1998. Folding nucleus: specific or multiple: insights from lattice models and experiments. *Fold. Des.* 3:R108–R111.
- Smith, L. J., C. M. Dobson, and W. F. van Gunsteren. 1999. Side-chain conformational disorder in a molten globule: molecular dynamics simulations of the A-state of human alpha-lactalbumin. *J. Mol. Biol.* 286:1567–1580.
- Steinbach, P. J., and B. R. Brooks. 1994. New spherical-cutoff methods for long-range forces in macromolecular simulation. *J. Comp. Chem.* 15:667–683.
- Ternstrom, T., U. Mayor, M. Akke, and M. Oliveberg. 1999. From snapshot to movie: phi analysis of protein folding transition states taken one step further. *Proc. Natl. Acad. Sci. U. S. A.* 96:14854–14859.
- Tirado-Rives, J., M. Orozco, and W. L. Jorgensen. 1997. Molecular dynamics simulations of the unfolding of barnase in water and 8 M aqueous urea. *Biochemistry.* 36:7313–7329.
- Tsai, J., M. Levitt, and D. Baker. 1999. Hierarchy of structure loss in MD simulations of src SH3 domain unfolding. *J. Mol. Biol.* 291:215–225.
- Walser, R., A. E. Mark, and W. F. van Gunsteren. 2000. On the temperature and pressure dependence of a range of properties of a type of water model commonly used in high-temperature protein unfolding simulations. *Biophys. J.* 78:2752–2760.
- Weng, X., H. Luecke, I. S. Song, D. S. Kang, S. H. Kim, and R. Huber. 1993. Crystal structure of human annexin I at 2.5 Å resolution. *Protein Sci.* 2:448–458.
- Wong, K. B., J. Clarke, C. J. Bond, J. L. Neira, S. M. Freund, A. R. Fersht, and V. Daggett. 2000. Towards a complete description of the structural and dynamic properties of the denatured state of barnase and the role of residual structure in folding. *J. Mol. Biol.* 296:1257–1282.