doi:10.1006/jmbi.2001.4552 available online at http://www.idealibrary.com on IDEAL® J. Mol. Biol. (2001) 309, 285–298

JMB

Role of Native Topology Investigated by Multiple Unfolding Simulations of Four SH3 Domains

Jörg Gsponer and Amedeo Caflisch*

Department of Biochemistry
University of Zürich
Winterthurerstrasse 190
CH-8057 Zürich, Switzerland

The relative importance of amino acid sequence and native topology in the unfolding process of two SH3 domains and two circular permutants was investigated by 120 molecular dynamics runs at 375 K for a total simulation time of 0.72 μs. The α-spectrin (aSH3) and src SH3 (sSH3) domains, which have the same topology and a sequence identity of only 34 %, show similar unfolding pathways. The disappearance of the three-stranded antiparallel β-sheet is the last unfolding event, in agreement with a large repertoire of kinetic data derived from point mutations as well as glycine insertions and disulfide crosslinks. Two alternative routes of β-sheet unfolding have emerged from the analysis of the trajectories. One is statistically preferred in aSH3 (n-src loop breaks before distal hairpin) and the inverse in sSH3. An elongation of the β2-β3 hairpin was observed during the unfolding of sSH3 at 375 K and in 300 K simulations started from the putative transition state of sSH3 in accord with unusual kinetic data for point mutations at the n-src loop. The change of connectivity in the permutants influenced the sequence of unfolding events mainly at the permutation site. Regions where the connectivity remained unaffected showed the same chronology of contact disappearance. Taken together with previous folding simulations of two designed three-stranded antiparallel β-sheet peptides, these results indicate that, at least for small β-sheet proteins, the folding mechanism is primarily defined by the native state topology, whilst specific interactions determine the statistically predominant folding route.

Keywords: molecular dynamics simulations; SH3 domain; Φ value analysis; native state topology; folding mechanism

*Corresponding author

Introduction

The protein folding problem remains a major challenge of biology despite the fact that experimental and simulation studies have led to considerable progress over the past few years (Dobson & Karplus, 1999). Both lines of investigation have provided interesting findings, amongst which the suggestion that for small proteins the folding mechanism and rates are determined mainly by the native state topology rather than the details of inter-residue interactions (reviewed by Baker, 2000). The study of the transition state for folding has attracted a lot of interest. Mutagenesis experiments on proteins with similar native topology have revealed that the structure of the folding transition state ensemble is conserved, despite differences in the sequence (Riddle *et al.*, 1999; Martinez & Serrano, 1999; Chiti *et al.*, 1999). On the other hand, differences in the folding pathway were found recently for proteins with the same native structure but only about 15 % sequence identity (McCallister *et al.*, 2000; Nishimura *et al.*, 2000). On the theoretical side, simple models based on native state topology have allowed the successful reproduction of essential features of transition-state structures (Galzitskaya & Finkelstein, 1999; Alm & Baker, 1999). A number of interesting questions are emerging from protein folding studies. Do proteins with the same native topology but different sequences have the same folding mechanism? Do multiple pathways exist for symmetric elements of structure? Are several alternative pathways equally consistent for a given topology?

The SH3 domains are a good system to study how the native topology influences the folding and unfolding mechanisms. They are globular proteins

Abbreviations used: MD, molecular dynamics; SAS, solvent-accessible surface.

E-mail address of the corresponding author: caflisch@bioc.unizh.ch

0022-2836/01/010285–14 $35.00/0

that consist of two antiparallel β-sheets (Figure 1). The X-ray structures of the α-spectrin SH3 domain (aSH3) (Musacchio *et al.*, 1992) and src SH3 domain (sSH3) (Xu *et al.*, 1997) show almost the same native conformation. A β-hairpin formed by the terminal segments is packed orthogonally on top of a three-stranded antiparallel β-sheet, also called central β-sheet. The kinetics and thermodynamics of folding of both proteins are well described by a two-state model (Viguera *et al.*, 1994; Grantcharova & Baker, 1997). Although the sequence identity between aSH3 and sSH3 is only 34 %, mutagenesis experiments suggest that major features of their transition state are similar (Martinez & Serrano, 1999; Riddle *et al.*, 1999; Grantcharova *et al.*, 2000). The central β-sheet, formed by strands 2-4, is believed to contain major parts of the folding nucleus in both domains (Martinez & Serrano, 1999; Riddle *et al.*, 1999; Grantcharova *et al.*, 2000). The crystal structures of two circular permutants of aSH3, S19-P20s (s19) (Figure 1(b)) and N47-D48s (n47) (Figure 1(c)), superimpose with the wild-type with a root-mean-square deviation of the C^{α} atoms (C^{α} rmsd) of less than 1 Å (Viguera *et al.*, 1996). In both permutants, the termini of aSH3 were connected with the same linker, Ser-Gly-Thr-Gly. The cleavage of the distal loop (between residues 47 and 48) induced no loss of hydrogen bonds in the crystal structure of n47. By contrast, the cut of the RT loop in s19 led to the loss of nine hydrogen bonds, five of them within the RT loop. The kinetic analysis showed that the s19 transition state ensemble differs less from that of aSH3 than n47 (Viguera *et al.*, 1996).

In this study, molecular dynamics (MD) simulations of four structurally related proteins, aSH3, sSH3, s19 and n47, were performed in an attempt to formulate an answer to the questions above. The simulations were carried out at 375 K with an implicit solvation model to sample a statistically significant number of unfolding events. aSH3 has been investigated by MD simulations to study its stability and dynamic behavior using various implicit solvent models and comparison with a 400 ps trajectory at 300 K in explicit solvent (van Aalten *et al.*, 1996). Recently, 30 unfolding runs of sSH3 at 498 K with explicit solvent were carried out to investigate the hierarchy of structure loss (Tsai *et al.*, 1999). The main differences between the present study and that by Tsai *et al.* are the lower temperature (375 K instead of 498 K) and the use of an implicit solvent. The latter is computationally more efficient than explicit solvent simulations, so a significant number of unfolding events could be sampled in a reasonable amount of time and at a lower temperature for sSH3 and for aSH3, s19 and n47.

The comparative analysis of the unfolding process of four proteins with the same structure but different sequence and/or connectivity reveals major similarities (for large parts of the pathway, including the late unfolding of the three-stranded β-sheet) and differences (at the permutation sites). In agreement with mutational analysis, the present simulation results indicate that the native topology determines the overall folding mechanism, which, for symmetrical substructures, allows for alternative pathways.

Results and Discussion

One control and 40 denaturation simulations of aSH3 and sSH3 were performed for a total of 492 ns (Table 1). In addition, one control and 20 denaturation simulations were carried out for s19 and n47 (252 ns). As in previous unfolding simulations (Lazaridis & Karplus, 1997, Ferrara *et al.*, 2000b), the fraction of native contacts, Q, was used as a progress variable to follow the average evolution of the system. To compare the sequence of unfolding events of the four proteins, only contacts between C^{α} atoms were defined as native contacts (see Materials and Methods; Figure 1(d)). At the denaturing condition of 375 K, 6 ns were sufficient to lead to major structural disorder or complete unfolding in all cases (Figure 2).

Table 1. Simulations performed

Protein	T (K)	Starting conformation[a]	Total number of simulations	Percentage of simulations with complete unfolding		
				0.05[b]	0.10[b]	0.20[b]
aSH3	300	1SHG	1			
	375	1SHG	40	93	98	100
	300	TS	10			
sSH3	300	1FMK	1			
	375	1FMK	40	60	68	83
	300	TS	5			
s19	300	1TUC	1			
	375	1TUC	20	80	85	95
n47	300	1TUD	1			
	375	1TUD	20	70	90	100

The length of the simulations was 6 ns except for the runs from the putative transition state, which lasted 10 ns.

[a] The pdb code is given for the simulations started from the X-ray conformation. TS stands for the putative transition state structures, i.e. the center of the clusters of the conformations with $0.5 < Q < 0.6$ sampled at 375 K.

[b] Fraction of native contacts used as a criterion for complete unfolding.

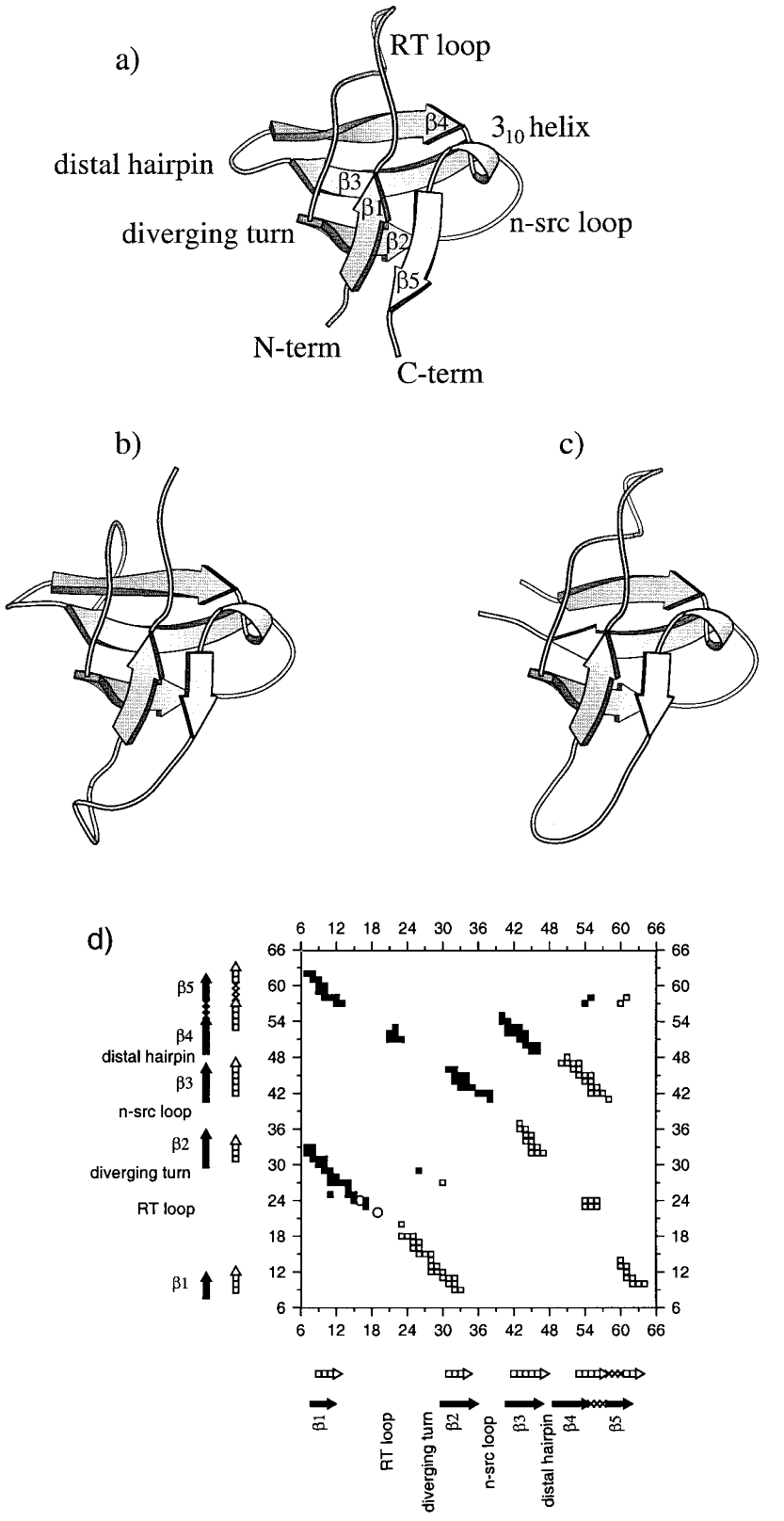

Figure 1. Structure comparison of the wild-type SH3 domain and two circular permutants. (a) Ribbon representation of the crystal structure of the α-spectrin SH3 domain (Musacchio *et al.*, 1992). (b) and (c) Ribbon representation of the crystal structure of the circular permutants S19-P20s and N47-D48s (Viguera *et al.*, 1996). (d) C^{α} contact map from the control run at 300 K. The contacts present for more than 4 ns of a total of 6 ns are shown. C^{α} contacts in aSH3 and the two permutants are shown above the diagonal (filled squares). Two contacts in the RT loop, which are represented with circles, were not present in s19. The contacts below the diagonal belong to sSH3 (open squares). The positions of the X-ray secondary structures are indicated along the axes. β-Strands are represented by arrows (black for aSH3 and the permutants and white for sSH3) and the 3_{10} helix by two crosses. The sequence alignment for aSH3 and sSH3 is shown at the bottom. Conserved residues are printed in bold. The ribbon diagrams were prepared using MOLSCRIPT (Kraulis, 1991).

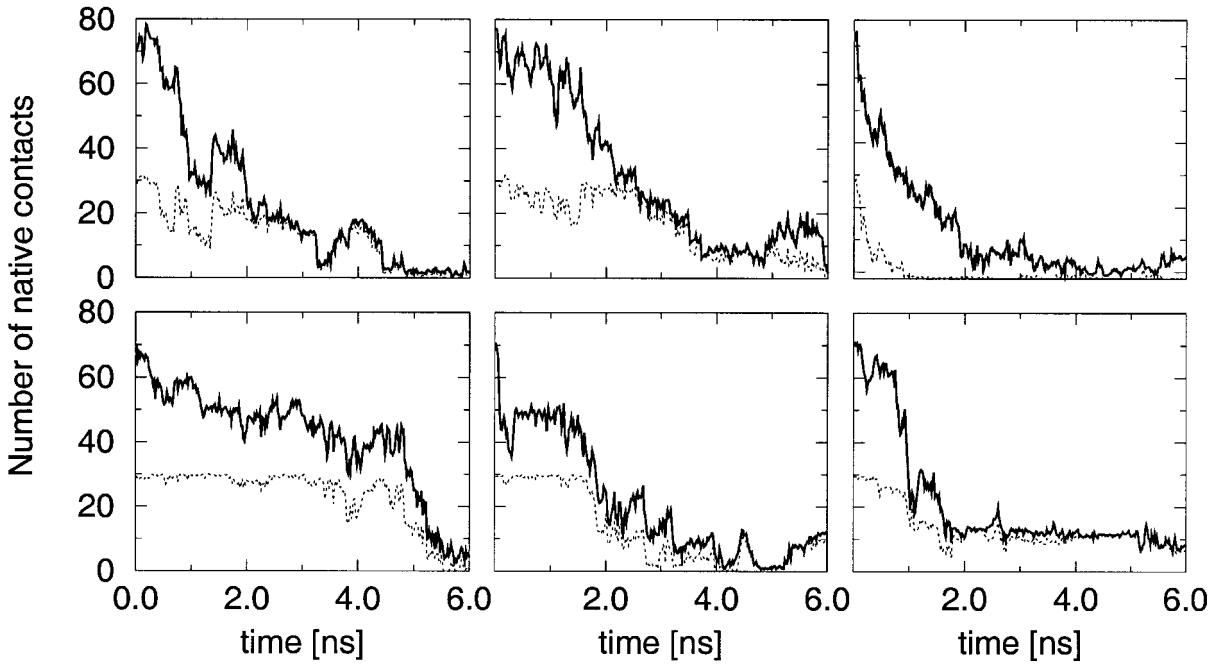

Figure 2. Evolution of the number of native contacts as a function of the simulation time for three aSH3 (top) and three sSH3 (bottom) unfolding trajectories. The continuous lines represent the total number of native contacts (Table 3) while the broken lines represent the native contacts of the central β-sheet.

Control runs

The structure of the native state was well maintained for all four proteins at 300 K. Neglecting the two N-terminal and two C-terminal residues, the C^{α} rmsd from the crystal structure averaged over the 6 ns is 1.77(±0.33) Å and 1.97(±0.21) Å for aSH3 and sSH3, respectively (Figure 3). The terminal residues displayed deviations of about 5 Å in both domains. The transiently higher C^{α} rmsd at 0.7, 3.5 and 5.9 ns in the aSH3 simulation was caused by increased deviations of the RT loop. This is consistent with explicit water MD simulations at room temperature, which indicated high flexibility for the RT loop and termini in both domains (van Aalten *et al.*, 1996; Tsai *et al.*, 1999). The fraction of native contacts, Q, averaged over the 6 ns at 300 K was 0.94 ± 0.25 and 0.94 ± 0.24 for aSH3 and sSH3, respectively.

In the two permutants, the fused loop, which has an unstructured conformation (Viguera *et al.*, 1996), as well as both termini deviated more from the crystal structure than the other parts of the protein. In particular, the N and C termini of s19, created by the cleavage of the RT loop, displayed important deviations (about 7 Å). This may be explained by the loss of five intra-RT loop hydrogen bonds in the crystal structure of s19. If one ignores the fused loop and the N-terminal and C-terminal residues (two in n47, as for the wild-type, and three in s19, due to the larger deviation of its termini), the average value of the C^{α} rmsd from the crystal structure was 1.75(±0.24) Å and 2.02(±0.38) Å for n47 and s19, respectively. The average value of Q over the 6 ns was 0.92 ± 0.27 and 0.90 ± 0.30 for n47 and s19, respectively.

Unfolding simulations

Simulations at high temperature can provide information on the kinetics and mechanisms of unfolding (Caflisch & Karplus, 1995; Lazaridis *et al.*, 1997; Wang *et al.*, 1999; Alonso & Daggett, 2000). In the case of multiple simulations for more than one protein sequence, interesting insights can be obtained on the relative kinetic stability. It is clear from Figure 2 that both aSH3 and sSH3, and especially the latter, unfold at different times as in a thermally activated process. The average C^{α} rmsd evolution is similar in the four proteins (Figure 3), although the deviation of sSH3 is smaller over the 6 ns and significantly smaller in the first half of the simulations. If the fraction of remaining C^{α} native contacts is used as a criterion to measure the degree of unfolding, significantly less sSH3 than aSH3 simulations reached the denatured state (Table 1). Differential scanning calorimetry and equilibrium fluorescence measurements indicate a slightly higher thermodynamic stability for sSH3 compared to aSH3 (Viguera *et al.*, 1994; Grantcharova & Baker, 1997). Hence, one would expect a slower unfolding of sSH3, if the unfolding time correlated with the increased thermostability. Nevertheless, kinetic experiments at 295 K have revealed a ten times higher unfolding rate for sSH3 than for aSH3 (Grantcharova & Baker, 1997; Viguera *et al.*, 1995). The kinetic data extracted from the simulations at 375 K are not

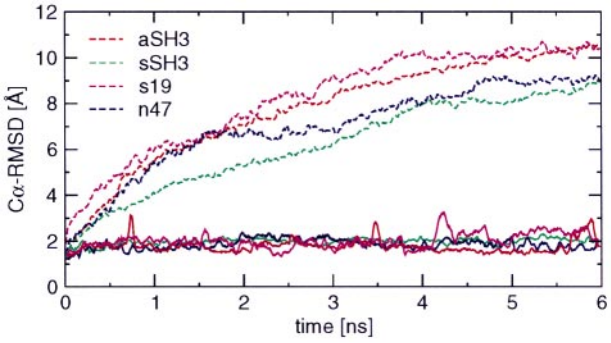

Figure 3. Evolution of the C^{α} rmsd (Å) from the crystal structure as a function of time in the simulations at 300 K (continuous lines) and 375 K (broken lines). The C^{α} rmsd values at high temperature represent averages over all simulations for each protein. The standard deviations are about 2 Å (not shown).

necessarily in disagreement with the experimental results at 295 K, since two proteins can have a significantly different dependence of their unfolding rate on the temperature, i.e. different slopes in the Arrhenius plot. At high temperature, the lines on the Arrhenius plot can cross and lead to an inversion of the tendency measured under standard conditions (Cavagnero *et al.*, 1998).

Unfolding mechanism

Although the unfolding trajectories of the four domains show some diversity, a statistically preferred chronology in the loss of native C^{α} contacts emerged from the simulations. Since the proteins unfolded at different times in the various simulations, the contact disappearance is reported in the following as a function of Q rather than time. The disappearance of native contacts is first compared between the topologically identical proteins aSH3 and sSH3, and then between aSH3 and its circular permutants.

aSH3 and sSH3

The average value of Q at the disappearance of a given native contact is shown in Figure 4(a) and representative pathways are illustrated in Figure 5. Contacts non-local in sequence (i.e. those far away from the diagonal in the plots of Figure 4) were the first to break apart in both proteins. In particular, the contacts between β1 and β5 as well as those between the RT loop and β4 were lost first. The latter can be considered as representatives of the hydrophobic core that connects the two orthogonal β-sheets of the SH3 domain. An early loss of the hydrophobic contacts in the core was observed in previous explicit water simulations performed by others with a different force-field (Tsai *et al.*, 1999). Intermediate kinetic stability was observed for the contacts between β1 and β2 and within the RT loop. The last contacts to fall apart in both proteins were in the central β-sheet composed of β2, β3, and β4. However, the most stable contacts were between strands β3 and β4 in aSH3, and between β2 and β3 in sSH3. The strong persistence of C^{α} contacts between β2 and β3 in sSH3 is in agreement with previous MD simulations in explicit water, in which hydrophobic contacts and hydrogen bonds between strands β2 and β3 were the most conserved (Tsai *et al.*, 1999).

The comparison of simulation data with experimental results can be only qualitative, since the exact localization of the transition state by MD simulations is not possible. The Φ values of both proteins suggest that the distal hairpin contributes to the transition state energetics, whereas the N and C termini consolidate very late in the folding process (Martinez & Serano, 1999; Riddle *et al.*, 1999). A similar finding was obtained by kinetic measurements of sSH3 mutants with a ten-glycine residue insertion or disulfide crosslink (Grantcharova *et al.*, 2000). These experiments investigate backbone ordering and supplement the Φ value analysis. The contacts between the distal hairpin are among the most stable in the simulations of both proteins, whereas those between the termini disappear first. This is consistent with the experimental data. The Φ values of residues lying in the n-src loop and the preceding strand β2 show discrepancies between the two proteins. This difference was explained partially by the fact that mutations at this positions probe the development of diverse structural elements in the two domains (Martinez & Serrano, 1999). Nevertheless, nine β2-β3-contacts that scan interactions between equivalent positions in the two domains disappeared at values of Q that were, on average, 0.34 higher in aSH3 than sSH3. If one excludes the nine β2-β3-contacts, there is a correlation of 0.81 for the average moment of disappearance (i.e. Q value at the rupture of a given contact averaged over all simulations of a given SH3 domain) of the remaining 53 contacts in common to aSH3 and sSH3 (Table 2). This is consistent with kinetic measurements. In fact, it was shown that positions for which Φ values have been determined in aSH3 and sSH3 have a correlation of 0.74 (Plaxco *et al.*, 2000). The simulations indicate that aSH3 and sSH3 show a similar sequence of events for the predominant part of the unfolding pathway. This may be related to the topological resemblance, because the sequence identity between the two proteins is only 34 %. However, the contacts between β2 and β3 disappeared earlier in aSH3 than in sSH3. Since the denaturation of the central β-sheet is the last unfolding event, it is probably less influenced by the native-state topology and more by the interactions between the side-chains (see also below).

aSH3 and s19

The order of disappearance of native contacts revealed one important difference; the contacts in

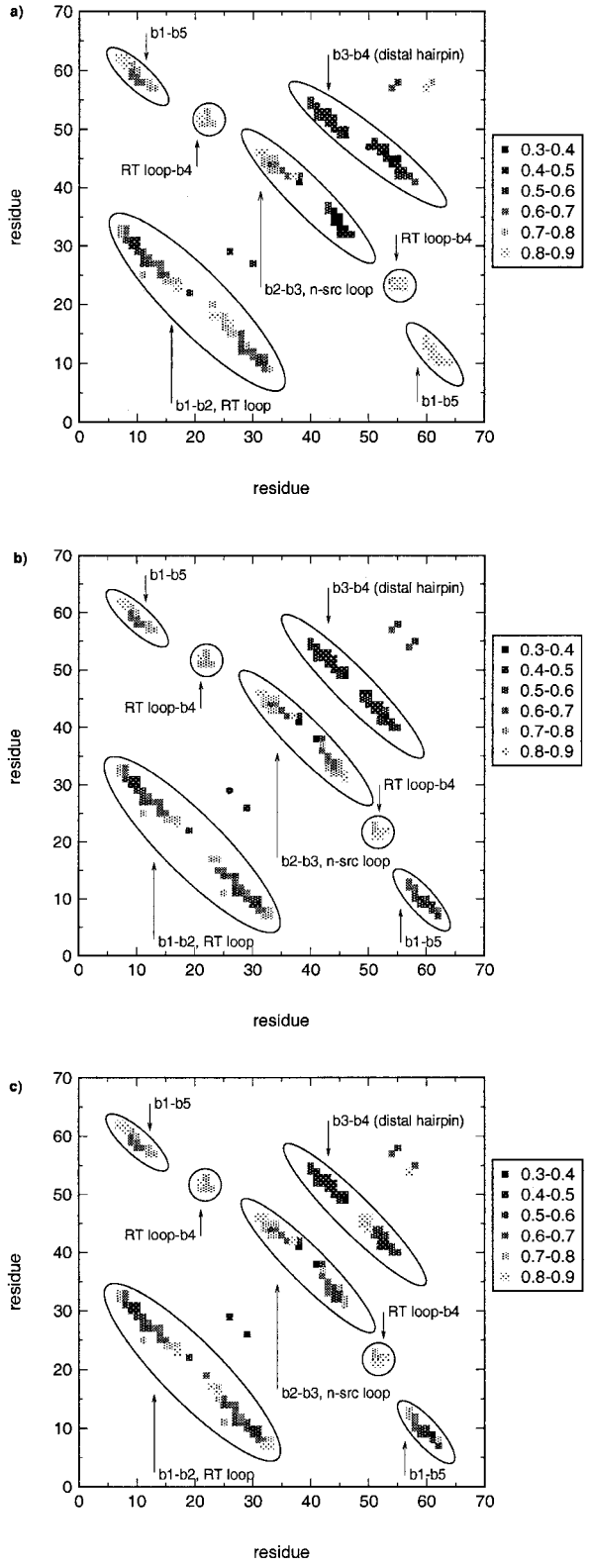

Figure 4. (a) Disappearance plot of aSH3 (above the diagonal) and sSH3 (below). The moment of first disappearance of each C^{α} contact, averaged over 40 simulations, is shown as a function of Q. The values of Q are indicated by different patterns according to the gradient on the right of the plot. Elements of secondary structure are highlighted. (b) Disappearance plot of aSH3 (above the diagonal) and s19 (below). The moment of first disappearance of the C^{α} contact of s19 is an average over the 20 unfolding simulations. (c) Disappearance plot of aSH3 (above the diagonal) and n47 (below). The moment of first disappearance of the C^{α} contact of n47 is an average over the 20 unfolding simulations.

the fused loop of s19 (between β1 and β5) fell apart much later than the corresponding contacts in aSH3 (Figures 4(b) and 5). This newly created hairpin, or at least parts of it, is expected to be fully folded in the transition state of s19, since Φ values close to 1 were determined for this region (Viguera *et al.*, 1995). As already seen above, contacts between the N and the C termini of aSH3, in contrast, are believed to contribute little to the transition state ensemble in terms of energy. For the remaining segments, the sequence of events was very much the same. In particular, the cleavage of the RT loop in s19 did not change the chronology of contact loss within this loop significantly. This may be related to the fact that in aSH3 the RT loop is highly flexible. The average moment of disappearance of the 67 contacts common to both proteins (excluding 11 contacts in the fused loop between β1 and β5) has a correlation of 0.94. This indicates that the circular permutant s19 has the same overall unfolding mechanism as aSH3.

aSH3 and n47

The creation of the new loop increased the persistence of the C^{α} contacts between β1 and β5 as for s19 (Figures 4(c) and 5), and the cleavage of the distal hairpin speeded the loss of the contacts close to the turn between β3 and β4. On the other hand, contacts between C^{α} atoms separated by more than two residues from the new termini disappeared as late as in aSH3. Furthermore, some changes were observed in the behavior of C^{α} contacts between β2 and β3. They persisted longer than the corresponding contacts in aSH3. Taking out 17 contacts that lay in the regions where the connectivity was changed (11 between β1 and β5 and six between β3 and β4) yielded a correlation of 0.83 in the average disappearance of the 63 contacts in common to aSH3 and n47. Hence, the overall unfolding mechanism of n47 is slightly more different from that of aSH3 than that of s19. These results are consistent with data from mutagenesis experiments, which indicate that the effect of the permutation on the transition state ensemble is more important in n47 than in s19 (Viguera *et al.*, 1996).

If one neglects the regions of the sequence adjacent to the permutation sites, the correlation in the native contact disappearance of aSH3, s19 and n47 is striking. The folding of the aSH3 domain is believed to follow the nucleation-condensation mechanism (Martinez *et al.*, 1998). When different contacts constitute the nucleus, as in the case of aSH3 and n47 (Viguera *et al.*, 1996), different folding mechanisms are expected. Therefore, one may

Table 2. Correlation of the native contact disappearance

Protein	aSH3	sSH3	s19	n47
aSH3	0.96[a]	0.81(53)[b]	0.94(67)[c]	0.83(63)[c]
sSH3		0.96	0.82(44)[d]	0.83(40)[d]
s19			0.90	0.84(71)[c]
n47				0.86

Fractional values represent correlation factors for the moment of contact disappearance. They are calculated by first averaging, over all simulations of a given protein, the value of Q at the rupture of a given contact. Then the correlation factor was calculated for the average Q values of the contacts in common.

[a] Values in the diagonal were determined by separating the 40(20) wild-type (permutant) simulations into two sets of 20(10) simulations each and calculating the correlation between the average values of Q at the moment of disappearance of native contacts in the two sets. A value close to 1.0 indicates that the average unfolding mechanism is the same in the two sets.

[b] Non-diagonal values close to 1.0 indicate that the two domains have a similar order of unfolding events. The value in parentheses is the number of common contacts excluding those between β2 and β3.

[c] Values in parentheses are the number of common contacts not adjacent to the permutation sites.

[d] Values in parentheses are the number of common contacts excluding those between β2 and β3, and those close to the permutation sites.

presume that the unfolding mechanism may show major discrepancies. However, in contrast to the classical nucleation model, where the native structure propagates in a stepwise manner from the nucleus, the nucleation-condensation model states that the formation of the nucleus is coupled with more general formation of structure (Fersht, 1999). The nucleus is simply the best-formed part at the transition state, but most of the protein chain is involved in forming it. Hence, due to the change in connectivity, another part may simply be the best formed at the transition state. The highly correlated

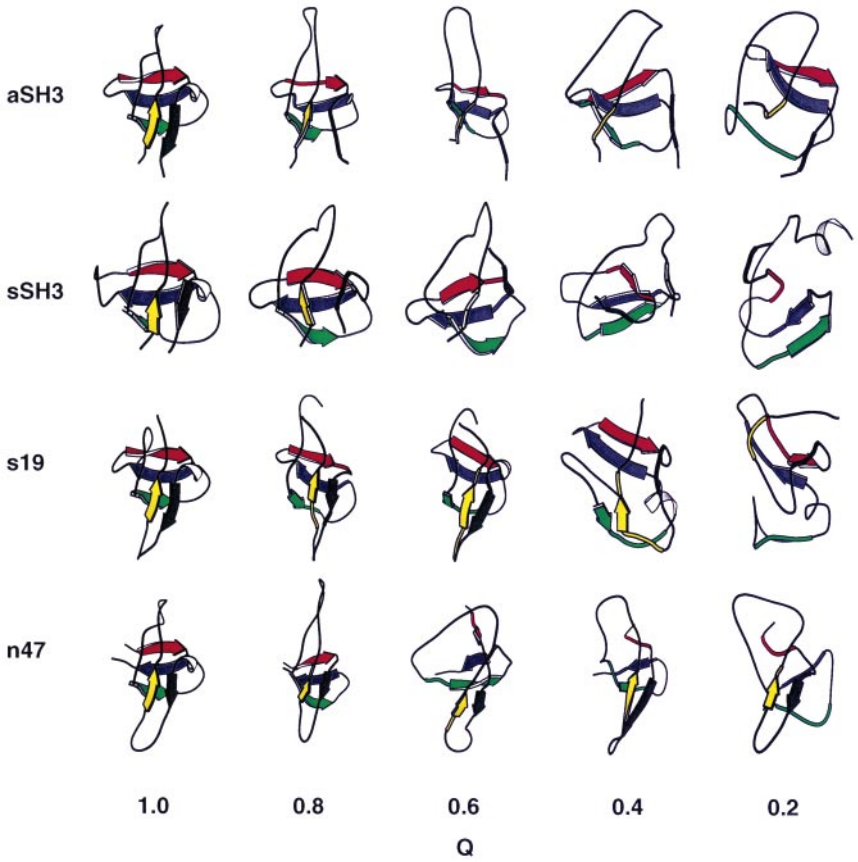

Figure 5. Ribbon diagrams of snapshots taken from representative denaturation trajectories. The five segments of sequence corresponding to the β-strands are shown in different colors: β1, yellow; β2, green; β3, blue; β4, red; β5, black. The secondary structure was determined with DSSP (Kabsch & Sander, 1983) and the plot was made with MOLSCRIPT (Kraulis, 1991).

disappearance of native contacts in the MD simulations indicates that the change in connectivity does not necessarily have an influence on the mechanism of folding and unfolding of structural elements that are not adjacent to the permutation sites.

Unfolding of the central β-sheet in aSH3 and sSH3

Mutagenesis experiments indicate that the three-stranded antiparallel β-sheet, or at least parts of it, could serve as folding nucleus in aSH3 and sSH3 (Martinez & Serrano, 1999; Riddle *et al.*, 1999; Grantcharova *et al.*, 2000). The statistically significant amount of unfolding trajectories allows an in-depth investigation of the kinetic stability and unfolding mechanism of the β-sheet. To monitor the unfolding process, useful progress variables are defined by taking into account the symmetry of the structure, as in previous studies (Ferrara & Caflisch, 2000, 2001; Hiltpold *et al.*, 2000). Q_2-3 and Q_3-4 are the fractions of the C^{α} contacts between β2 and β3 and between β3 and β4, respectively (Table 3). Q_2-3-4 is defined as the fraction of the C^{α} contacts in the central β-sheet.

Two-dimensional projections into the progress variables Q_non2-3-4 and Q_2-3-4 (Figure 6(a),(b)), as well as Figure 4, indicate that the triple-stranded β-sheet unfolds late, on average. In 22 of 40 aSH3 simulations, most contacts outside of the three-stranded β-sheet were lost before the β-sheet began to unfold. Contacts started to break first between β2-β3 or β3-β4 in only seven runs. In the remaining 11 runs, the loss of contacts was concerted. Even clearer was the behavior for sSH3. In 36 simulations, contacts between β2, β3 and β4 began to break when the majority of the other contacts had already disappeared. In 13 of these, the simulation time was not long enough for a complete unfolding of the three-stranded β-sheet with an average of Q_2-3-4 = 0.8 for the final 200 ps. In contrast to aSH3, contacts between β2, β3 and β4 were never lost first. A concerted contact loss was observed in only four runs. Hence, the central β-sheet was the kinetically most stable element in both proteins, but particularly in sSH3. These results are consistent with recent experimental data, which indicate that the entire three-stranded β-sheet remains intact at the unfolding transition state of sSH3, whereas the N and C termini are dissociated and the RT loop is opened (Grantcharova *et al.*, 2000). Moreover, NMR experiments indicate that preferentially native-like topology of the central β-sheet is sampled in the denatured state ensemble of aSH3 (Kortemme *et al.*, 2000).

The projection into the Q_2-3Q_3-4 plane indicates a different sequence of events for the unfolding of the triple stranded β-sheet in the two domains (Figure 6(c)-(f)). In 25 of the 40 aSH3 runs, all of the contacts between the strands β2 and β3 were lost before those between β3 and β4 started to disappear. The reverse was observed in only ten runs. In sSH3, by contrast, unfolding started by the loss of contacts between β3 and β4. Eighteen of the 27 simulations in which the three-stranded β-sheet unfolded, showed a complete loss of contacts between β3 and β4 first. In only seven runs, contacts disappeared between β2 and β3 first. A concerted loss of contacts between the three strands was observed only rarely, i.e. in five aSH3 and two sSH3 simulations.

Recently, reversible folding simulations of a designed three-stranded antiparallel β-sheet peptide of 20 residues have shown the existence of two alternative routes for folding, and one of the two is statistically predominant (Ferrara & Caflisch, 2000). In the main pathway, the β-hairpin consisting of the middle and C-terminal strand is formed first. This is followed by the consolidation of the N-terminal strand onto the preformed β-hairpin. The other pathway consists of the inverse sequence of events, i.e. the N-terminal hairpin forms first and the C-terminal strand coalesces afterwards. Furthermore, unfolding was shown to be the reverse of folding at both 330 K and 360 K (Ferrara & Caflisch, 2000). More recent simulations of two three-stranded β-sheet peptides with a sequence identity of 15 % indicate that the amino acid sequence determines which of the two routes is statistically preferred (Ferrara & Caflisch, 2001). In the SH3 domain, β2, β3 and β4 do not form a model three-stranded antiparallel β-sheet, since β2 and β3 are connected by a five (seven) residue loop in aSH3 (sSH3). Yet, the simulation results for the antiparallel β-sheet in the SH3 domains correspond to the behavior observed for the synthetic 20 residue peptides, at least during unfolding. The

Table 3. Native C^{α} contacts

Protein	Total[a]	β2-β3-β4[b]	β2-β3[c]	β3-β4[d]
aSH3	80	32	15	17
sSH3[e]	72(62)	30(24)	12(9)	18(15)
s19	78	32	15	17
n47	80	32	15	17

[a] Total number of native C^{α} contacts.
[b] Number of C^{α} contacts in the central β-sheet, β2, β3 and β4.
[c] Number of C^{α} contacts between β2 and β3.
[d] Number of C^{α} contacts between β3 and β4.
[e] Values in parenthesis are the contacts in common with aSH3.

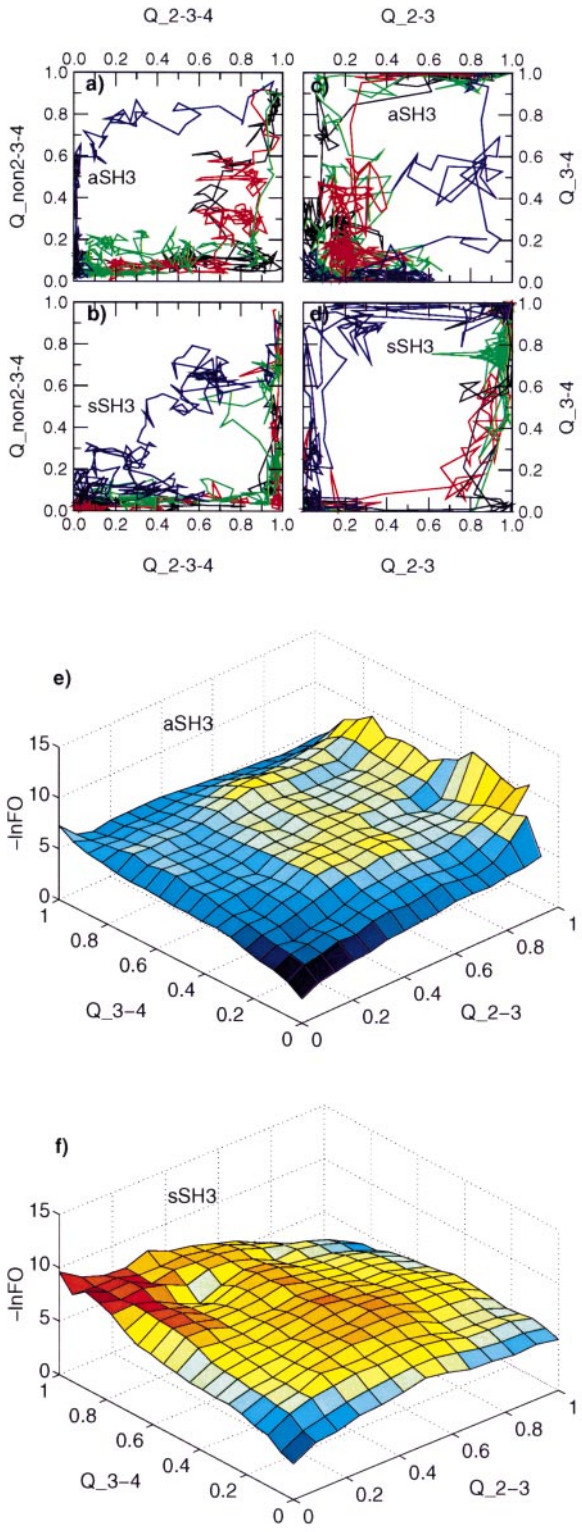

Figure 6. Projection into the *Q*_non2-3-4*Q*_2-3-4-plane and the *Q*_2-3*Q*_3-4-plane of (a) and (c) four aSH3 and (b) and (d) four sSH3 trajectories at 375 K. The red, black and green trajectories are representatives of the statistically preferred unfolding pathways, whereas the blue trajectories represent the less populated pathways. The blue trajectory in (a) is one of the seven aSH3 simulations, where contacts were first lost between β2, β3 and β4. In (b), the concerted loss of contacts within and outside the central β-sheet is represented with the blue trajectory. In (a)-(d), each point corresponds to the average over 20 ps (ten conformations). (e) and (f) Negative logarithm of the frequency of occurrence (FO) as a function of the fraction of the native contacts in the strands β2 and β3 (*Q*_2-3), and β3 and β4 (*Q*_3-4). The value of each bin was computed as $-\ln(N_{m,m}/N_{\mathrm{tot}})$, where $N_{n,m}$ denotes the number of conformations with $n(m)$ contacts formed between strands β2 and β3 (β3 and β4). The error was estimated by separating the 40 simulations into two sets of 20 simulations each. The average error is 0.27 and 0.57 for aSH3 and sSH3, respectively. The maximal error is 2.96 for the bin (14,0) in aSH3 and 2.20 for the bin (3,17) in sSH3.

two routes are populated differently in aSH3 and sSH3. One sequence of unfolding events of the three-stranded β-sheet is statistically predominant in aSH3, whereas the inverse one is the most populated in sSH3. The different unfolding behavior appears to be related to differences in specific interactions in the central β-sheet (the 27 residues in β2-β3-β4 have a sequence identity of only 30 %) rather than topology, since the latter is essentially the same in the two domains. Recent kinetic data indicate that for proteins with symmetric structures, several alternative folding routes may be consistent with the native state topology (McCallister *et al.*, 2000). The preferred route is likely to involve the formation of the substructures with the lowest free energy; hence, it is determined by the sequence.

Elongation of the β2-β3-hairpin during the unfolding of sSH3

Figure 7 depicts the fraction of time during which backbone polar groups were involved in hydrogen bonds between β2 and β3 in the control run and unfolding simulations. All of the backbone hydrogen bonds of the X-ray structure were conserved at 300 K, apart from the hydrogen bond between the oxygen atom of Arg31 and the nitrogen atom of Leu48, which was broken early and then reformed temporarily in the 300 K run of sSH3. The hydrogen bonding pattern was conserved before the unfolding of aSH3 at 375 K. By contrast, during most of the 375 K simulations of sSH3, an elongation of strand β2 was observed as a consequence of new hydrogen bonds between Asn36 and Leu44, Thr38 and Trp42 as well as between Glu39 and Trp42. Moreover, the hydrogen bond between Arg31 and Leu48 was formed more frequently.

A cluster analysis was performed to analyze the frequency of conformations with an elongation of strand β2. For both domains, the initial ensemble contained every tenth structure (12,000 conformations). The first and second cluster of sSH3 incorporate 17.8 % and 4.5 % of all structures, respectively; the first ten clusters 30.9 %. A DSSP analysis showed that the center of the first cluster is the native state conformation. The center of the

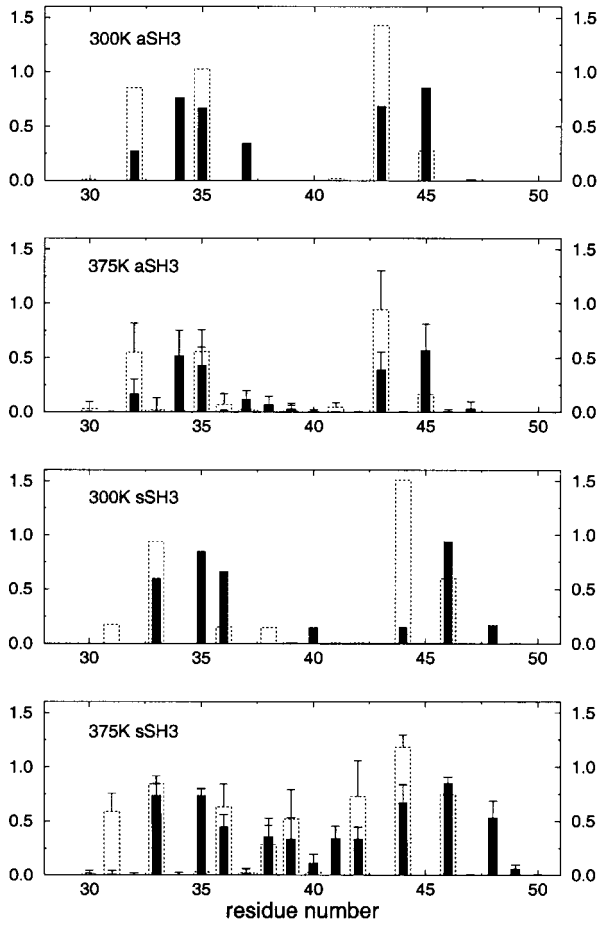

Figure 7. Backbone hydrogen bonds between β2 and β3. Depicted is the average over all simulations of the fraction time (*y*-axes), during which backbone oxygens act as hydrogen bond acceptors (stippled bars) and backbone nitrogens as donors (black bars). For the unfolding simulations, only the time intervals previous to the unfolding of the three-stranded β-sheet were taken into account. Values larger than 1.0 are due to bifurcated hydrogen bonds.

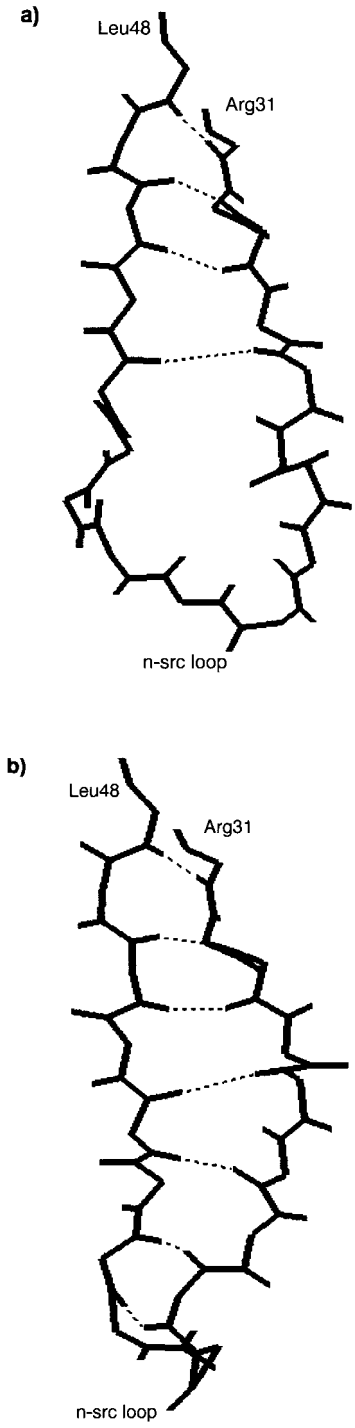

Figure 8. Backbone representation of (a) the n-src loop of the X-ray structure, and (b) the center structure of the second most populated cluster derived from the cluster analysis of the 40 sSH3 unfolding simulations at 375 K.

second cluster has an intact three-stranded β-sheet with an elongated strand β2 (Arg31-Thr38) followed by a turn formed by residues Glu39-Asp41 (Figure 8). Moreover, six of the first ten clusters have a center structure with an elongated β2-β3-hairpin. These results suggest the presence of metastable states with a tight hairpin connecting an elongated strand β2 with β3. This is consistent with the Φ values greater than 1 of the n-src loop residues Asn37 and Gly40 in sSH3 (Riddle *et al.*, 1999). Riddle *et al.* proposed, that the residues in the n-src loop might be ordered at the transition state in a non-native conformation, i.e. as a tight hairpin, rather than a disordered loop. Tsai *et al.* did not mention the elongation of β2-β3 in the analysis of their explicit water MD simulations at 498 K (Tsai *et al.*, 1999). Recent computer simulations with a cubic lattice model indicate that the folding nucleus can contain specific interactions not present in the native state (Li *et al.*, 2000). Such interactions in the transition state would give rise to Φ values that are negative or larger than unity, as in the n-src loop of sSH3.

In aSH3, the first cluster contains 9.7 % of all conformations and the native state is its center structure. The first ten clusters incorporate 16.1 % of all structures and none of them had an elongation of β2. An in-depth comparison with the experimental data is difficult, since the Φ value of only one residue in the n-src loop of aSH3 was determined (Martinez & Serrano, 1999). However, Ser36 has a Φ value of 0.25 in aSH3, whereas its structurally equivalent position in sSH3, Asn37, has a Φ value of 3.0. The low Φ value of Ser36 appears to be in agreement with the absence of an elongation of strand β2 during the unfolding of aSH3.

Since a longer and more stable non-native n-src hairpin in the transition state of sSH3 might be one of the reasons for the different sequence of unfolding events of the central β-sheet in aSH3 and sSH3, an additional set of simulations at 300 K were started from putative transition state conformations. To define the transition state ensemble, the 4552 partially unfolded sSH3 conformations with $0.5 < Q < 0.6$ obtained in the 375 K runs were clustered. The choice of $0.5 < Q < 0.6$ is consistent with the fact that in most 375 K runs Q shows the steepest slope in this interval (Figure 2), and with the definition used by others (Tsai *et al.*, 1999). The center structure of the first ten clusters represents 41 % of the conformations with $0.5 < Q < 0.6$. They were submitted to 10 ns simulations at 300 K. Eight of the ten cluster representatives have an elongated strand β2 that remained stable during the 10 ns; the time series for one of these is given in the middle of Figure 9. Two clusters have no extension of strand β2, but during the simulation additional β-hairpin hydrogen bonds are formed between β2 and β3 (Figure 9(b), right). The elongation of strand β2 is concomitant with the formation of van der Waals contacts between the apolar part of the Glu39 and Trp42 side-chains. These contacts are absent in the native state and the 300 K runs from the X-ray structure of sSH3 (Figure 9(c), left). As suggested by the Φ value analysis (Riddle *et al.*, 1999), Ile34 appears to be strained in the native conformation because of overpacking of the hydrophobic core, but not in the transition state. The simulation results are consistent with the experiments, since in the 300 K runs from the putative transition state, Ile34 and Trp43 have less interactions with Asn37 and Phe10, respectively, than in the control run from the X-ray structure (Figure 9, bottom). Apart from a less dense packing of the hydrophobic core, mainly due to the unfolding of Phe10 and the relaxation of the Ile34 and Trp43 overpacking, no major change in interactions between residues on β2 and β3 could be observed.

In aSH3, the putative transition state ensemble ($0.5 < Q < 0.6$) used for the cluster analysis contained 5754 conformations. The first ten clusters

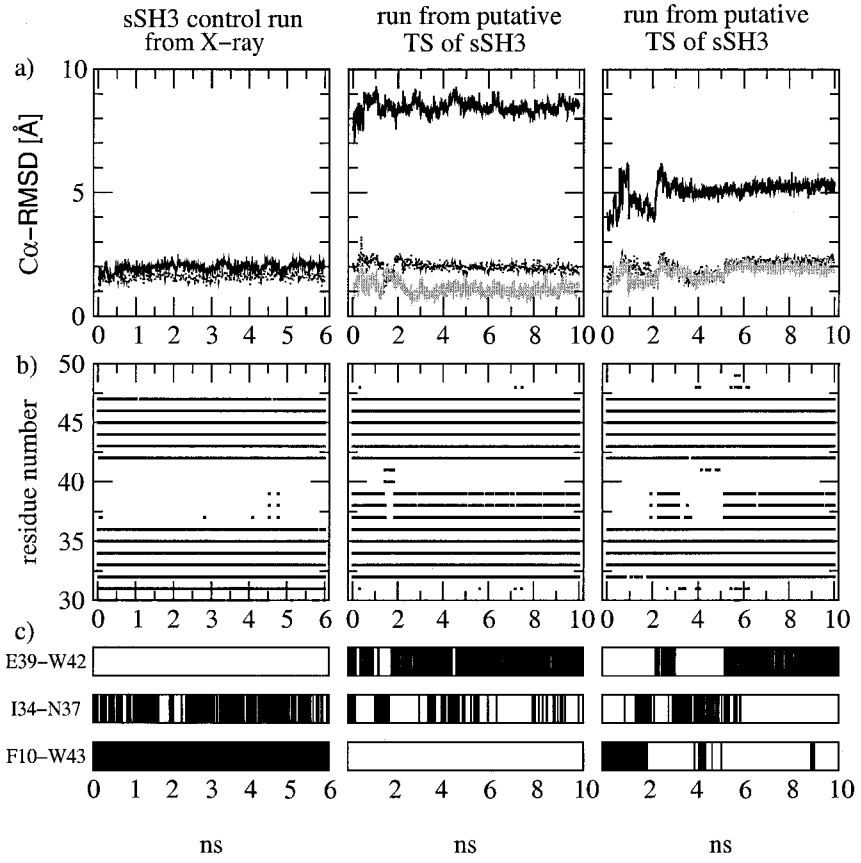

Figure 9. Time series of the sSH3 control run at 300 K (left plots) and two simulations started from putative transition state conformations (middle and right plots). (a) Evolution of the all C^{α} rmsd (continuous lines) and the central β-sheet C^{α} rmsd (Å) (broken lines) from the crystal structure. Evolution of the central β-sheet C^{α} rmsd (Å) from the starting structure (grey lines). (b) β-strand content in the segment of sequence between Glu30 and Thr50 determined with DSSP (Kabsch & Sander, 1983). (c) Contacts between side-chains are represented by vertical dashes when the centers of geometry of the side-chains are within 7.0 Å.

incorporate 36 % of all structures. Only one cluster representative shows an elongation of strand β2, whilst in five clusters strand β2 was completely unfolded. In the remaining four representatives, strand β2 had the same number of backbone hydrogen bonds as in the native state (two clusters) or less (two clusters). The five cluster representatives with a partially or completely folded strand β2 were used as initial structures for simulations at 300 K. No elongation of β2 was observed; moreover, in the only starting conformation with more backbone hydrogen bonds between β2 and β3 than in the X-ray structure of aSH3, the hydrogen bonds disappeared in the first 4 ns (data not shown).

The starting conformations used in the ''transition state'' simulations represent only a putative transition state ensemble, since the exact localization along the MD trajectories at 375 K is not possible (Du *et al.*, 1998). Nevertheless, the results confirm that the formation of non-native backbone hydrogen bonds between β2 and β3 is frequent in sSH3 and absent in aSH3. Furthermore, the extension of β2 was seen during the ten sSH3 simulations at 300 K started from a partially unfolded structure but not in the one initiated from the native state, which indicates that a less packed hydrophobic core might facilitate the elongation of the β-hairpin β2-β3.

Conclusions

The unfolding process of four SH3 domains was simulated by multiple MD runs at 375 K to investigate the influence of differences in sequence and connectivity on the order of events. Although the transition state is difficult to localize in the MD trajectories, the simulation results are in good qualitative agreement with mutagenesis experiments, and this allows us to analyze the sequence of events with more confidence. aSH3 and sSH3 have almost the same topology and significant similarities for large parts of their unfolding pathway. This is consistent with the correlation of 0.74 for the Φ values of equivalent positions in the two proteins (Plaxco *et al.*, 2000). In agreement with experimental data, the loss of interactions between the termini and opening of the RT loop are the first unfolding events. The disappearance of the three-stranded antiparallel β-sheet is the last unfolding event and proceeds first by an almost complete rupture of contacts in one of the two hairpins. This allows for two main routes of unfolding of the central β-sheet. The order of hairpin rupture in the statistically predominant unfolding pathway of aSH3 is the inverse of that in the most populated pathway of sSH3. Recent experimental data indicate that, for proteins with symmetric structures, more than one route to the folded state may be equally consistent with the native state topology (McCallister *et al.*, 2000). Moreover, previous folding simulations of two designed three-stranded antiparallel β-sheet peptides showed the existence of alternative folding pathways (Ferrara & Caflisch, 2000, 2001). The 375 K simulations of sSH3 indicate the presence of a non-native tight hairpin connecting an elongated strand β2 with β3 that is not observed for aSH3. Simulations at 300 K started from ten putative transition state conformations of sSH3 indicate that the elongated β-hairpin is stable over a 10 ns time-scale. The non-native hairpin β2-β3 in the transition state of sSH3 might be one of the reasons for the different predominant routes of unfolding of the central β-sheet between aSH3 and sSH3.

The change of connectivity in the permutants influenced the sequence of unfolding events mainly at the permutation site. The same order of contact disappearance was observed for regions where the connectivity remained unaffected. Moreover, the cleavage of the already highly flexible RT loop in s19 had little influence on the loss of contacts within this loop. Taken together with mutagenesis data, the simulation results indicate that the order of (un)folding events of small β-sheet proteins is determined mainly by the native topology. However, symmetric substructures, e.g. a three-stranded β-sheet, show multiple alternative routes of (un)folding and the statistically preferred pathway seems to be determined by the amino acid sequence.

Materials and Methods

Model

All proteins were modelled by explicitly considering all heavy atoms and the hydrogens bound to nitrogen or oxygen atoms using the CHARMM PARAM19 force-field (Brooks *et al.*, 1983). A mean-field model of the aqueous solvent (Ferrara *et al.*, 2000c) was employed to overcome the problem of time-consuming explicit water simulations. The CHARMM PARAM19 default cutoff for long-range interactions was used, i.e. a shift function (Brooks *et al.*, 1983) was employed with a cutoff at 7.5 Å for both the electrostatic and the van der Waals terms. To approximate the screening effects of the electrostatic interactions, the model uses a distance-dependent dielectric function, $\varepsilon(r) = 2r$ (Ferrara *et al.*, 2000c), and neutralized side-chains (Lazaridis & Karplus, 1997). This leaves the short-range hydrogen bonding interactions almost unaltered while screening the electrostatic interaction of partial charges at larger distances. All other solvation effects are approximated with a model based on the solvent-accessible surface (SAS) (Eisenberg & McLachlan, 1986). The mean solvation term is given by:

$$V_{\mathrm{solv}}(r) = \sum_{i=1}^{N} \sigma_i A_i(r) \qquad (1)$$

for a protein having N heavy atoms with Cartesian coordinates $r = (r_1, \ldots r_N)$. $A_i(r)$ is the solvent-accessible surface area of atom i, computed by an approximate analytical expression (Hasel *et al.*, 1988) and using a 1.4 Å probe radius. The model contains only two different surface tension parameters; one for carbon and sulfur atoms ($\sigma_{C,S} = 0.012$ kcal mol^{-1} Å^{-2}) and one for nitrogen

and oxygen atoms ($\sigma_{N,O} = -0.060$ kcal mol^{-1} Å^{-2}) (Ferrara *et al.*, 2000a). Previous folding and unfolding simulations of CI2 with this implicit solvation model were in good agreement with experimental results and explicit water simulations (Ferrara *et al.*, 2000a). More recently, the SAS model was used to simulate the reversible folding of two three-stranded antiparallel β-sheet peptides of 20 residues each (Ferrara & Caflisch, 2000, 2001) and an α-helix of 31 residues (Hiltpold *et al.*, 2000). Furthermore, the same force-field and solvation model were used to demonstrate the non-Arrhenius behavior of the temperature-dependence of the folding rate (Ferrara *et al.*, 2000c).

Simulations

The crystal structures of the SH3 domain of chicken α-spectrin (1SHG) (Musacchio *et al.*, 1992), its circular permutants S19-P20s (1TUC) and N47-D48s (1TUD) (Viguera *et al.*, 1996), as well as the src SH3 domain (1FMK) (Xu *et al.*, 1997) were minimized using a distance-dependent dielectric function by 400 steps of steepest descent and 600 steps of conjugate gradient to relieve local strain. The minimized structures had a C^α rmsd from the X-ray conformation of 0.58, 0.57, 0.57 and 0.44 Å, respectively, and were used as starting points for the MD simulations. Constant-temperature MD simulations were performed at either 300 K or 375 K (Table 1) with a weak coupling (Berendsen *et al.*, 1984). The SHAKE algorithm (Ryckaert *et al.*, 1977) was used to fix the length of the covalent bonds where hydrogen atoms are involved; this allows us to use an integration time step of 2 fs. The non-bonded interactions were updated every ten steps and structures were saved every 2 ps. The only difference for the multiple runs at 375 K was the seed for the random number generator used for the initial assignment of the velocities.

Analysis methods

The C^α contacts are defined as pairs of C^α atoms of residues not adjacent in sequence (i.e. residue pairs i, j with $j > i + 2$) having a distance smaller than 6.5 Å for more than two-thirds of the 300 K simulations (Sheinerman & Brooks, 1998). Eighty C^α contacts were found in aSH3 and n47 (Table 3). While n47 has the same contacts as aSH3, the cleavage of the RT loop in s19 reduced the number of C^α contacts to 78. For sSH3, 72 C^α contacts were found, 62 of which are in common with aSH3. For the unfolding simulations, a contact was considered broken if the distance between the C^α atoms was greater than 7.0 Å for more than ten consecutive integration steps (20 ps).

The method for the cluster analysis is based on structural similarity (Ferrara & Caflisch, 2000). The C^α rmsd is calculated for each pair of structures after optimal superposition. The number of neighbors is then calculated for each conformation using a C^α rmsd cutoff of 3.5 Å. The conformation with the highest number of neighbors is considered as the center (or representative) of the first cluster. All the neighbors of this conformation are removed from the ensemble of conformations. The center of the second cluster is then determined in the same way as for the first cluster and this procedure is repeated until each structure is assigned to a cluster.

The hydrogen bond criterion consisted of a distance between the amide hydrogen atom and the oxygen atom less than 2.6 Å and the N-H$\cdots$O angle larger than 120°.

Acknowledgments

We thank Ph. Ferrara and S. Fioravanti for helpful discussions, and N. Budin for computer support. We thank an anonymous referee for suggesting that we run the simulations at 300 K from the putative transition state conformations. This work was supported, in part, by the Swiss National Science Foundation (grant no. 31-53604.98 to A.C.). J.G. is a fellow of the MD-PhD program (grant no. 3236-057617).

References

Alm, E. & Baker, D. (1999). Prediction of protein-folding mechanisms from free-energy landscapes derived from native structures. *Proc. Natl Acad. Sci. USA,* **96**, 11305-11310.

Alonso, D. O. & Daggett, V. (2000). Staphylococcal protein A: unfolding pathways, unfolded states, and differences between the B and E domains. *Proc. Natl Acad. Sci. USA,* **97**, 133-138.

Baker, D. (2000). A surprising simplicity to protein folding. *Nature,* **405**, 39-42.

Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. & Haak, J. R. (1984). Molecular dynamics with coupling to an external bath. *J. Phys. Chem.* **81**, 3684-3690.

Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. & Karplus, M. (1983). CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **4**, 187-217.

Caflisch, A. & Karplus, M. (1995). Acid and thermal denaturation of barnase investigated by molecular dynamics simulations. *J. Mol. Biol.* **252**, 672-708.

Cavagnero, S., Debe, D. A., Zhou, Z. H., Adams, M. W. & Chan, S. I. (1998). Kinetic role of electrostatic interactions in the unfolding of hyperthermophilic and mesophilic rubredoxins. *Biochemistry,* **37**, 3369-3376.

Chiti, F., Taddei, N., White, P. M., Bucciantini, M., Magherini, F., Stefani, M. & Dobson, C. M. (1999). Mutational analysis of acylphosphatase suggests the importance of topology and contact order in protein folding. *Nature Struct. Biol.* **6**, 1005-1009.

Dobson, C. M. & Karplus, M. (1999). The fundamentals of protein folding: bringing together theory and experiment. *Curr. Opin. Struct. Biol.* **9**, 92-101.

Du, R., Pande, V. S., Grosberg, A. Y., Tanaka, T. & Shakhnovich, E. I. (1998). On the transition coordinate for protein folding. *J. Chem. Phys.* **108**, 334-350.

Eisenberg, D. & McLachlan, A. D. (1986). Solvation energy in protein folding and binding. *Nature,* **319**, 199-203.

Ferrara, P. & Caflisch, A. (2000). Folding simulations of a three-stranded antiparallel β-sheet peptide. *Proc. Natl Acad. Sci. USA,* **97**, 10780-10785.

Ferrara, P. & Caflisch, A. (2001). Native topology or specific interactions: what is more important for protein folding? *J. Mol. Biol.* **306**, 837-850.

Ferrara, P., Apostolakis, J. & Caflisch, A. (2000a). Computer simulations of protein folding by targeted molecular dynamics. *Proteins: Struct. Funct. Genet.* **39**, 252-260.

Ferrara, P., Apostolakis, J. & Caflisch, A. (2000b). Targeted molecular dynamics simulations of protein unfolding. *J. Phys. Chem. ser. B,* **104**, 4511-4518.
Ferrara, P., Apostolakis, J. & Caflisch, A. (2000c). Thermodynamics and kinetics of folding of two model peptides investigated by molecular dynamics simulations. *J. Phys. Chem. ser. B,* **104**, 5000-5010.
Fersht, A. (1999). *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*, 3rd edit., Freeman, New York.
Galzitskaya, O. V. & Finkelstein, A. V. (1999). A theoretical search for folding/unfolding nuclei in three-dimensional protein structures. *Proc. Natl Acad. Sci. USA,* **96**, 11299-11304.
Grantcharova, V. P. & Baker, D. (1997). Folding dynamics of the src SH3 domain. *Biochemistry,* **36**, 15685-15692.
Grantcharova, V. P., Riddle, D. S. & Baker, D. (2000). Long-range order in the src SH3 folding transition state. *Proc. Natl Acad. Sci. USA,* **97**, 7084-7089.
Hasel, W., Hendrickson, T. F. & Still, W. C. (1988). A rapid approximation to the solvent accessible surface areas of atoms. *Tetrahedron Comput. Methodol.* **1**, 103-116.
Hiltpold, A., Ferrara, P., Gsponer, J. & Caflisch, A. (2000). Free energy surface of the helical peptide Y(MEARA)6. *J. Phys. Chem. ser. B,* **104**, 10080-10086.
Kabsch, W. & Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers,* **22**, 2577-2637.
Kortemme, T., Kelly, M. J., Kay, L. E., Forman-Kay, J. & Serrano, L. (2000). Similarities between the spectrin SH3 domain denatured state and its folding transition state. *J. Mol. Biol.* **297**, 1217-29.
Kraulis, P. (1991). MOLSCRIPT, a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946-950.
Lazaridis, T. & Karplus, M. (1997). ''New view'' of protein folding reconciled with the old through multiple unfolding simulations. *Science,* **278**, 1928-1931.
Lazaridis, T., Lee, I. & Karplus, M. (1997). Dynamics and unfolding pathways of a hyperthermophilic and a mesophilic rubredoxin. *Protein Sci.* **6**, 2589-2605.
Li, L., Mirny, L. A. & Shakhnovich, E. I. (2000). Kinetics, thermodynamics and evolution of non-native interactions in a protein folding nucleus. *Nature Struct. Biol.* **7**, 336-42.
Martinez, J. C., Pisabarro, M. T. & Serrano, L. (1998). Obligatory steps in protein folding and the conformational diversity of the transition state. *Nature Struct. Biol.* **5**, 721-729.
Martinez, J. C. & Serrano, L. (1999). The folding transition state between SH3 domains is conformationally restricted and evolutionarily conserved. *Nature Struct. Biol.* **6**, 1010-1016.
McCallister, E. L., Alm, E. & Baker, D. (2000). Critical role of beta-hairpin formation in protein G folding. *Nature Struct. Biol.* **7**, 669-673.
Musacchio, A., Noble, M., Pauptit, R., Wierenga, R. & Saraste, M. (1992). Crystal structure of a Src-homology 3 (SH3) domain. *Nature,* **359**, 851-855.
Nishimura, C., Prytulla, S., Jane Dyson, H. & Wright, P. E. (2000). Conservation of folding pathways in evolutionarily distant globin sequences. *Nature Struct. Biol.* **7**, 679-686.
Plaxco, K. W., Larson, S., Ruczinski, I., Riddle, D. S., Thayer, E. C., Buchwitz, B., Davidson, A. R. & Baker, D. (2000). Evolutionary conservation in protein folding kinetics. *J. Mol. Biol.* **298**, 303-312.
Riddle, D. S., Grantcharova, V. P., Santiago, J. V., Alm, E., Ruczinski, I. & Baker, D. (1999). Experiment and theory highlight role of native state topology in SH3 folding. *Nature Struct. Biol.* **6**, 1016-1024.
Ryckaert, J. P., Ciccotti, G. & Berendsen, H. J. C. (1977). Numerical integration of the cartesian equation of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comp. Phys.* **23**, 327-341.
Sheinerman, F. B. & Brooks, C. L. (1998). Calculations on folding of segment B1 of streptococcal protein G. *J. Mol. Biol.* **278**, 439-456.
Tsai, J., Levitt, M. & Baker, D. (1999). Hierarchy of structure loss in MD simulations of src SH3 domain unfolding. *J. Mol. Biol.* **291**, 215-225.
van Aalten, D. M., Amadei, A., Bywater, R., Findlay, J. B., Berendsen, H. J., Sander, C. & Stouten, P. F. (1996). A comparison of structural and dynamic properties of different simulation methods applied to SH3. *Biophys. J.* **70**, 684-692.
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.
Viguera, A. R., Blanco, F. J. & Serrano, L. (1995). The order of secondary structure elements does not determine the structure of a protein but does affect its folding kinetics. *J. Mol. Biol.* **247**, 670-681.
Viguera, A. R., Serrano, L. & Wilmanns, M. (1996). Different folding transition states may result in the same native structure. *Nature Struct. Biol.* **3**, 874-880.
Wang, L., Duan, Y., Shortle, R., Imperiali, B. & Kollman, P. A. (1999). Study of the stability and unfolding mechanism of BBA1 by molecular dynamics simulations at different temperatures. *Protein Sci.* **8**, 1292-1304.
Xu, W., Harrison, S. C. & Eck, M. J. (1997). Three-dimensional structure of the tyrosine kinase c-Src. *Nature,* **385**, 595-602.

Edited by B. Honig

(Received 4 September 2000; received in revised form 28 December 2001; accepted 15 February 2001)