WILEY InterScience® DISCOVER SOMETHING GREAT

PROTEINS STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

Molecular dynamics simulations from putative transition states of α-spectrin SH3 domain

Xavier Periole,[1]* Michele Vendruscolo,[2] and Alan E. Mark[1,3,4]*

[1] Department of Biophysical Chemistry, Groningen Biomolecular Sciences and Biotechnology Institute (GBB), University of Groningen, 9747 AG Groningen, The Netherlands

[2] Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, United Kingdom

[3] School of Molecular and Microbial Sciences, University of Queensland, St Lucia, 4072, Queensland, Australia

[4] Institute for Molecular Biosciences, University of Queensland, St Lucia, 4072, Queensland, Australia

ABSTRACT

A series of molecular dynamics simulations in explicit solvent were started from nine structural models of the transition state of the SH3 domain of α-spectrin, which were generated by Lindorff-Larsen et al. (Nat Struct Mol Biol 2004;11:443–449) using molecular dynamics simulations in which experimental Φ-values were incorporated as restraints. Two of the nine models were simulated 10 times for 200 ns and the remaining models simulated two times for 200 ns. Complete folding was observed in one case, while in the other simulations partial folding or unfolding events were observed, which were characterized by a regularization of elements of secondary structure. These results are consistent with recent experimental evidence that the folding of SH3 domains involves low populated intermediate states.

Proteins 2007; 69:536–550.

Key words: molecular dynamics; transition state; protein folding; phi-value; Pfold.

INTRODUCTION

Understanding of the process of protein folding is one of the grand challenges in molecular biology. Ever since it was shown that the information contained in the sequence of amino acids is sufficient for a protein to find the structure of its native state,[1] experimentalists and theoreticians have tried to understand the mechanisms of this important biological process.[2–5]

Much effort has been focused on proteins that undergo two-state folding.[5] The primary advantage of two-state proteins is the lack of detectable intermediate states, so their folding process can be considered as involving a transition from a broad ensemble of configurations representing the unfolded state to a narrow ensemble of configurations making up the native state, via a specific transition state. The analysis of two-state proteins greatly simplifies the interpretation of experiments designed to elucidate the mechanism of folding. The Φ-value analysis,[6] in which the effects of specific amino acid substitutions on folding kinetics and equilibria are measured,[7] has been widely used to obtain structural information regarding the nature of the transition state as Φ-values reflect the degree to which the environments of specific residues are native-like in the transition state. By assuming that Φ-values correlate with the proportion of native contacts in the transition state,[8–13] Φ-values can also be used as restraints in computer simulations. This approach has been used by several groups to propose models for the structures of the transition state ensembles (TSEs) of a range of proteins.[10,12,14] However, as independent experimental information about transition states is difficult to obtain, it has been problematic to verify whether a TSE generated in this manner is in fact representative of the true TSE. Studies in which the TSE generated using a set of experimental Φ-values is used to predict the results of further Φ-value measurements,[10,15] and studies showing that the contact order in the TSE correlate with protein folding rates[16] are supporting the use of this approach. A further validation has been provided recently by the demonstration that the results of double-mutant cycle experiments could be predicted from the knowledge of a set of structures representing the TSE of barnase.[17]

In an alternative theoretical approach, a particular structure or set of structures is identified as belonging to the TSE by calculating the probability of folding by generating trajectories that are started from the proposed structures.[18] Since the TSE corre-

Grant sponsor: European Community Training and Mobility of research "Protein (mis)folding"; Grant number: HPRN-CT-2002-00241.
*Correspondence to: Xavier Periole, Groningen Biomolecular Sciences and Biotechnology Institute (GBB), Department of Biophysical Chemistry, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands.
E-mail: x.periole@rug.nl or Alan E. Mark, School of Molecular and Microbial Sciences, University of Queensland, St. Lucia, 4072, Queensland, Australia. E-mail: a.mark@uq.edu.au
Received 28 November 2006; Revised 22 February 2007; Accepted 7 March 2007
Published online 10 July 2007 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.21491

sponds to the point of highest free energy along the reaction coordinate of the folding process, a TSE structure should have an equal probability to fold or to unfold, and therefore its folding probability, Pfold, should be equal to 0.5.[18] Several studies have used this type of approach to validate TSEs. Gsponer and Caflisch[13] generated putative TSE conformations and estimated the Pfold values of six structures using an implicit model to account for solvation effects. They found that up to 200 ns simulations were necessary to discriminate between folding or unfolding behavior. Shakhnovich and coworkers determined Pfold values in simulations using a Go potential[19,20] and observed that not all conformations satisfying experimental Φ-values were part of the TSE on the free energy landscape of the model that they considered.[20] More recently, however, Wolynes and coworkers concluded that the use of Φ-values as restraints represents an effective strategy for the accurate determination of transition state structures.[21] Go potentials, in which native interactions are considered more favorable than nonnative ones, were used both by Shakhnovich and coworkers and Wolynes and coworkers, and they have the advantage of being computationally very convenient. There has been, however, some debate about whether the free energy landscapes[22] and the folding kinetics[23] of Go models resemble the true ones faithfully enough to provide good estimates of the Pfold values, which are known to be extremely sensitive to changes in conformations and energetics.[23] Within this context, the use of explicit solvent models in Pfold calculations has been analyzed recently by Rhee and Pande,[23] who showed that it can represent, at least in the case of the 23-residue mini-protein BBA5, a promising approach for a reliable estimation of Pfold. Even with simplified models the evaluation of reliable Pfold values is computationally extremely demanding. Several methods have been proposed to reduce this demand by alternative determinations of the Pfold values of conformations sampled during reversible folding simulations.[24–26] For example Rao *et al.*[24] showed that the Pfold of a conformation can be approximated by the probability that structurally similar conformations (clusters) fold during a reversible folding simulation. Such a procedure was suggested to be equivalent to a Pfold calculation for each individual conformation of the cluster obtained through multiple simulations but much less computational demanding.

In this work, extensive atomistic molecular dynamics (MD) simulations in explicit solvent are used to examine a set of nine structures selected from a model of the TSE ensemble of the α-spectrin SH3 domain, which was generated by Lindorff-Larsen *et al.*[27] using experimental Φ-values as restraints. Two of the structures were simulated for 200 ns for 10 times while the remaining ones for two times. Taken together, the results discussed are derived from a total of 6.8 μs of MD. The experimentally determined native conformation was also simulated for 100 ns for comparison. In the following we examine a range of structural properties, including the solvent accessible surface area, the radius of gyration, the deviation of the structure from the native configuration (global, local, and per residue), and ratio of native contacts.

MODEL AND METHODS

Starting structures for the simulations

The native state

The native structure of the SH3 domain of α-spectrin (PDB code 1BK2[28]) was used. The numbering of the residues is as in Ref. 27.

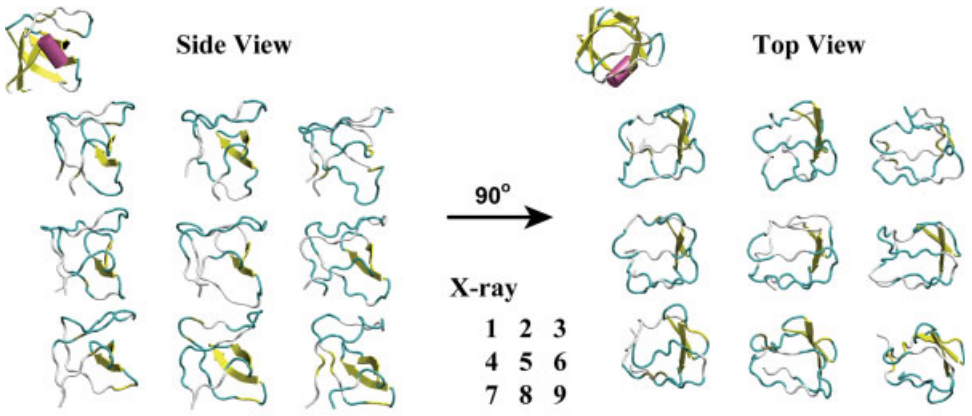
The transition state ensemble

We considered the structures representing the TSE of α-spectrin SH3 generated by Lindorff-Larsen *et al.*[27] In brief, this TSE was generated using MD simulations[12] in which experimental Φ-values, Φ^{exp}, were used as restraints by adding a term in the force field to penalize the difference between the experimental Φ-values and those calculated during the simulations, Φ^{calc}. The Φ^{exp} values represent the ratio of the destabilization of the transition state (TS), $\Delta\Delta G_i^{TS-U}$, compared to that of the native state (N), $\Delta\Delta G_i^{N-U}$, due to a mutation *i*,[7] and were interpreted as a measure of ratio of native contacts, Φ^{calc}, present in the TSE.[10,12,27]

$$\Phi_i^{exp} = \frac{\Delta\Delta G_i^{TS-U}}{\Delta\Delta G_i^{N-U}} \quad (1)$$

The unfolded state (U) is used as reference. The Φ^{calc} values were computed for each residue in a given conformation as the ratio of native contacts[8,10]: $\Phi_i^{calc} = Q_i^{conf}/Q_i^N$, where Q_i^{conf} designate the number of contacts of the residue *i* in a conformation and Q_i^N the number of contacts of the same residue *i* in the native state. A specific contact between two residues was considered to exist if the two residues were separated by at least two positions along the sequence and the distance between their Cα atoms was less than 8.5 Å.[10] A contact was considered native if observed to exist for more than 80% of the time during the 100 ns simulation of the native state. Φ^{calc} were determined for each conformation of the MD simulations.

Nine transition state conformations (TSC) were extracted from the original set of 500 structures (the set $\Theta = 500$ K, see Ref. 27 for details) and will be referred as TSC-*X* with $X = 1$ to 9. In Figure 1, a cartoon representation of their structures and some of their general properties are shown. The root mean square distance (RMSD) from the native state of the Cα atoms, which ranges from 4.1 to 5.1 Å, the radius of gyration, R_{gyr}, and the solvent accessible surface area, SASA, indicate



Data	TSC-1	TSC-2	TSC-3	TSC-4	TSC-5	TSC-6	TSC-7	TSC-8	TSC-9	X-ray
rmsd (Å)	4.3	4.5	4.1	4.4	4.2	4.8	4.1	5.1	5.1	-
SASA ($x10^2.Å^2$)	204.0	206.2	202.2	206.9	209.1	204.3	206.3	206.3	203.4	185.0
S.S.	13(8)	14(8)	8(4)	18(10)	10(8)	12(8)	19(12)	21(14)	25(12)	35(26)
Rgyr (Å)	11.2	11.1	11.1	11.2	11.2	11.2	11.1	11.4	11.2	10.2

Figure 1

Ribbon diagrams of the native and transition state conformations analyzed in this work. The topology of the native state (top left corner) is shown both in top and in side view, together with the nine TSCs (see inset for TSC numbers). Values for the all-Cα rmsd (rmsd), the solvent accessible surface area (SASA), the number of residues involved in a secondary structure element (S.S.) and in a β-strand (in parenthesis), and the radius of gyration (R_{gyr}) are reported. The last column corresponds to the X-ray structure (1bk2). [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

that the nine TSCs differ significantly from the X-ray structure of the native state although the overall native topology is maintained (Fig. 1). With the exception of β3 and β4 few secondary elements are systematically present in the TSCs.

MD simulations

All MD simulations were performed using the GROMACS-3.1 program package.[29,30] The GROMOS force field 43a1[31] was used to describe the protein and the SPC[32] water model for the solvent. Each system (see below) was solvated in a cubic box and simulated at 1 atm and 300 K. Both the temperature and pressure were maintained close to their target values using the Berendsen[33] weak coupling algorithm ($\tau_T = 0.1$ ps and $\tau_P = 1$ ps). A twin-range cut-off (1.0–1.4 nm) was used for the nonbonded interactions. Interactions within the short-range cutoff were evaluated every time step (2 fs), whereas interactions within the longer range cutoff were evaluated every 10 steps together with the pair-list. To correct for the truncation of electrostatic interactions beyond the long-range cutoff, a Reaction-Field[34] correction was applied ($\epsilon = 78$). Bond lengths were constrained using the LINCS[35] algorithm for the protein and the SETTLE[36] algorithm for the water.

Native state

The native state was solvated in a box of SPC water leaving a minimum of 1.0 nm from the protein to the edge of the box and simulated for 100 ns. This trajectory was used to build a set of observables that were used to characterize the native fold. Most of them were averaged over the full length of the simulation.

Transition state

Each TSC was solvated in a box of SPC water molecules leaving 1.0 nm from the protein to the edge of the box. After energy minimization, a 50 ps simulation was performed with position restraints applied on all heavy atoms of the protein. A second 50 ps simulation was then performed with restraints only on the heavy atoms of the backbone. This was done to ensure that the starting configuration was preserved during solvation. The system was then simulated without structural restraints. Each TSC was simulated twice for 200 ns. The simulations

DOI 10.1002/prot

were labeled TSC-*X*-1 and TSC-*X*-2, respectively, with $X = 1$ to 9.

To further evaluate the folding propensities from the TSCs, eight additional runs of 200 ns each were performed for two conformations, TSC-4 and TSC-5. These structures were chosen as they had the greatest apparent tendency to fold or unfold in the first two simulations (see the Results section). New random velocities were assigned to the system for each run. These simulations will be referred to as TSC-4-1 to TSC-4-10 and TSC-5-1 to TSC-5-10.

Analysis of the simulations

To evaluate the extent to which a TSC acquired native-like properties, the following properties were monitored.

Solvent accessible surface area (SASA): The SASA was computed using the double cube lattice algorithm[37] with a probe radius of 1.4 Å. The values of the starting conformation and the average over the 100 ns simulation of the native state are reported for the native fold. In the case of the TSCs the values for the starting model, and for the average from 199 to 200 ns for each simulation, are reported.

Secondary structure: The time evolution of the secondary structure (as defined by DSSP[38]) of the TSC runs was monitored to evaluate the formation of elements of secondary structure during the simulation. The number of residues involved in specific elements of secondary structure (α-helix, β-sheet, β-strand) was also determined and compared to that observed in the simulation of the native fold.

Global root mean square deviation (RMSD): The RMSD of the Cα atoms (excluding the N and C termini: residues 1 and 57) of the TSCs from the X-ray model was calculated to quantify the overall deviation of the structure of each TSC from the native fold.

Net-9 and Net-6: The time series of the RMSD of residues that were identified as key amino acids for folding or stability of the domain[27] were also monitored. Two subsets were considered. Net-9 contains the nine major hydrophobic core residues, labeled I to IX (residues 4, 6, 18, 20, 26, 28, 39, 48, and 53, respectively). Net-6 contained a subset of net-9 (III to VIII). It has been suggested that residues in net-6 form an important contact network for the folding of the protein.[27] These six residues are distributed in the RT-loop, β2, β3, and β4. The three residues omitted (I, II, and IX) belong to β1 and β5, two interacting β-strands.

Local RMSD: To determine the extent of formation of local, native-like, tertiary structure element, subsets of Cα atoms were defined based on the secondary structure elements (SSE) present in the native fold. The subsets of residues used are the following: all-Cα: 2–56, RT-Loop: 8–20, β-strand 1 (β1): 2–6, β-strand 2 (β2): 25–30, β-strand 3 (β3): 36–40, β-strand 4 (β4): 45–49, β-strand 5 (β5): 53–55, β-sheet contains β1, β2, β3, β4, and β5. Pairs of consecutive (often interacting) strands were also used: β1β2, β2β3, β3β4, β4β5, and β1β5. Note that β4 and β5 are not in contact but are linked by a small 3_{10} helix. The latter was too small to be defined as an independent subset.

RMSD per residue: The time series of the RMSD per residue were used to monitor the evolution of the protein structure at the residue level. The RMSD of each Cα was calculated after a least square fit of the conformation to the X-ray model using all Cα atoms.

RESULTS AND DISCUSSION

The native fold

We monitored various structural parameters derived from the 100 ns simulation of the native fold (Fig. 2). Specifically, Figure 2(a) shows the secondary structure as a function of time, and Figure 2(b) shows the RMSD of

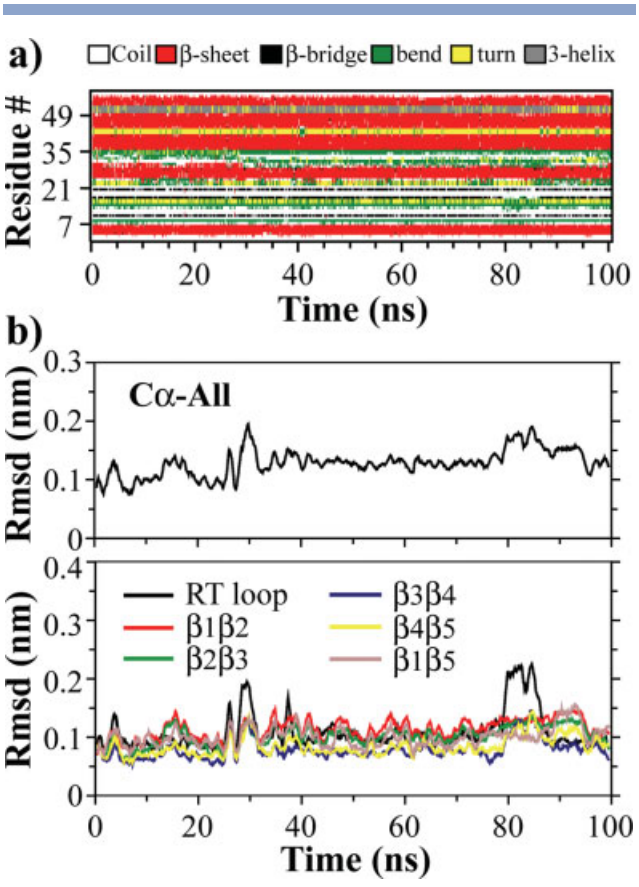


Figure 2

*Native state simulation of the α-spectrin SH3 domain. Time series of: (**a**) the secondary structure, and (**b**) the root mean square deviation of subsets of Cα (see Method section for definition of the subsets) after a least square fit of the all-Cα subset on the X-ray structure. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]*

Table I
TSC-1 to TSC-9 after 200 ns Simulation

TSC no.		1		2		3		4		5		6		7		8		9		<>	Native
R_{gyr}	Init	11.2		11.1		11.1		11.2		11.2		11.2		11.1		11.4		11.2		11.2	10.3
	Run 1, 2	10.7	10.3	10.6	10.2	10.6	10.5	10.7	10.2	10.6	10.9	10.3	10.7	10.3	10.6	10.3	10.4	10.3	10.5	10.5	
SASA	Init	204.0		206.2		202.2		206.9		209.1		204.3		206.3		206.3		203.4		205.4	186.4
	Run 1, 2	187.3	185.8	182.4	183.7	186.6	183.3	189.6	185.4	183.5	186.7	183.8	182.4	184.1	185.1	186.9	185.2	186.7	188.9	185.4	
S.S.	Init	13/8		14/8		8/4		18/10		10/8		12/8		19/12		21/14		25/12		16/9	32/24
	Run 1, 2	27/22	22/13	17/6	18/9	27/10	24/16	31/23	31/24	29/18	36/29	25/14	20/12	27/21	15/8	31/29	27/22	27/20	28/22	26/18	

The table lists the radius of gyration (R_{gyr}, Å), the solvent accessible surface area (SASA, 10^2 Å^2), and the number of residues involved in a secondary structure element (S.S); first the total number is given and then the number of residues involved in a β-strand. Init is the value for a TSC in the TSE. Run 1, 2 are, respectively, the average values for the last ns of the 200 ns simulation of TSC-*X*-1 and TSC-*X*-2 where $X = 1$ to 9. The values reported in the Native column are averages over the 100 ns of the native fold simulation. See Method section for more details.

all Cα atoms as well as different subsets of Cα atoms from the native state. From these results, the native fold can be considered as very stable in the force field. In fact only the RT loop shows variations greater than 2 Å at any stage during the whole simulation.

The transition state configurations

To examine the structural evolution of the trajectories started from the TSCs various properties including the radius of gyration (R_{gyr}), the solvent accessible surface area (SASA), changes in secondary structure, the global, local RMSD, RMSD per residue, and Φ^{calc}-values were analyzed as a function of time. In presenting the results, four examples have been selected that illustrate the alternative behaviors observed during the simulations. These include TSC-4-2, which shows complete folding to a native like stable state, TSC-6-1, which shows partial folding, TSC-5-2, which unfolds forming an excess of secondary structure, and TSC-2-1, which unfolds with little formation of secondary structure.

R_{gyr} and SASA

The values of R_{gyr} and SASA for the TSCs at the beginning and the end of the 200 ns simulations, together with the value obtained for the native fold, are reported in Table I. In all cases there was a systematic decrease in the R_{gyr} relative to the initial structures with values approaching that of the native fold. There was also a systematic decrease in the SASA in the TSCs, with values again approaching that of the native structure. These results suggest that the structures obtained by Lindorff-Larsen *et al.*[27] are more expanded than those favored by the force field used in this study.

Secondary structure

The TSCs contain few of the SSEs found in the native fold (Fig. 1). Only two β-strands, β3 and β4, are clearly identifiable in the initial structures (β-strands: yellow). They form a two-stranded β-sheet (yellow arrow) in all structures except TSC-3. Nevertheless, the topology of the native fold is clearly visible in all of the structures shown in Figure 1.

The total number of residues involved in a SSE, together with the number of residues involved in a β-strand, are reported in Table I. There is a net increase of both these quantities compared to the initial structure during the simulations. This finding illustrates that there is a strong tendency for the SSEs to become more ordered during the simulations.

The time evolution of the secondary structure is shown in Figure 3 for representative TSC examples and can be compared with that of the native structure shown in Figure 2(a). The general regularization of the secondary structure is clearly evident in TSC-4-2, TSC-6-1, and

DOI 10.1002/prot

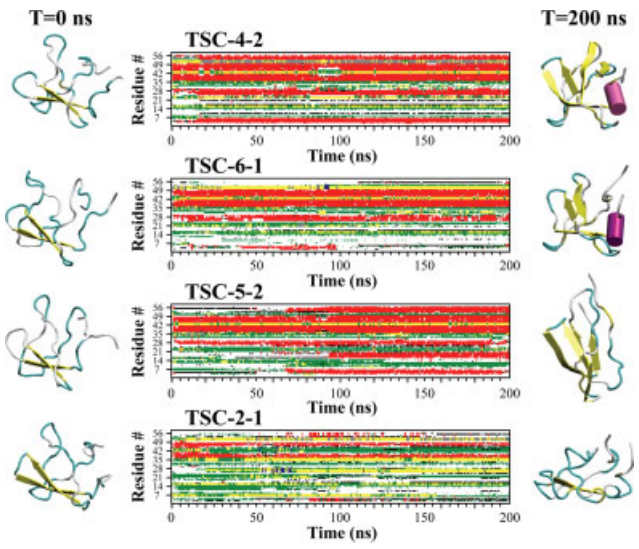


Figure 3

Time series representing the evolution of the secondary structure elements. A selection of four scenarios is presented: complete folding (TSC-4-2), partial folding (TSC-6-1), unfolding with excess of secondary structure formation (TSC-5-2), and unfolding with little secondary structure formed (TSC-2-1). The initial (T = 0 ns) and final (T = 200 ns) structures of the TSCs are shown as a cartoon representation. The color code is the same used in Figure 1.

TSC-5-2. The β-strands β3 and β4 remained intact in almost all of the simulations, but were lost in certain cases, e.g. in TSC-2-1. No other native SSE formed consistently. β1, β2, β5, and the 3_{10} helix, which lies between β4 and β5, appeared transiently in most simulations. In most cases the folding behaviors of β1 and β5 were correlated. α-helical segments also appeared transiently but in most cases did not remain stable. Note that in the case of TSC-4-2 the missing native SSEs formed simultaneously after 20 ns and remained stable for the following 180 ns of the simulation.

Although native-like elements of secondary structure formed during the simulations, the overall shape of the molecule underwent significant distortions in some cases namely TSC-2-1, TSC-3-1, and TSC-5-2. The native-like topology, which was proposed to be characteristic of the TSE,[27] was lost in these three simulations. For example, in the case of TSC-5-2, shown in Figure 3, β5 (C-term) lost all contact with β1 (N-term) instead interacting with RT Loop in a completely nonnative configuration.

Analysis of the RMSD from the native state

The deformation of the TSCs relative to the native fold was quantified in terms of the RMSD of the Cα atoms (all-Cα subset) from the X-ray structure. Only in one simulation (TSC-4-2) was the native state reached [Fig. 4(a)]. In this case, after 20 ns the all-Cα RMSD from the X-ray model was less than 1.5 Å. The native fold was maintained for the rest of the simulation. In three other cases the native fold was approached, with RMSD values of 1.2, 1.9, and 2.4 Å for TSC-4-1, TSC-3-2, and TSC-6-2, respectively. Importantly, the adoption of the native-like conformation in these three cases occurred at the later stages of the simulation, namely after 180 ns, but was not stabilized. In contrast in both simulations of TSC-5 (TSC-5-1 and TSC-5-2) the molecule rapidly unfolded with the RMSD reaching values exceeding 7.0 Å. In the remaining simulations the RMSD values ranged from 3 to 5 Å. Thus, in the majority of cases it was not possible to discriminate between folded and unfolded states based on the global RMSD. Interestingly, the average values of the all-Cα RMSD of the initial and the final conformations are similar. These results indicate that during the simulations the TSC become as much native-like as unfolded, as is also apparent from the comparison of the initial and final values of each TSC in Table II.

As an alternative to the global RMSD we also examined two sets of residues that had been proposed to form key interactions in the native and in the transition states: net-9, which contains the nine major hydrophobic core residues, and net-6, a subset of net-9. These two sets of residues have been suggested to be linked into an important contact network for the folding of the protein.[27] The RMSD values of the net-9 and net-6 subsets are reported in Table II for the starting and final conformations. As these residues are well dispersed within the folded state of the protein, the RMSD of these residues is sensitive to the overall shape of the protein. Indeed the RMSD of net-9 residues is correlated to that of all Cα atoms throughout the simulations (data not shown). The discrimination between folded and unfolded states based on the RMSD of net-9 and net-6 is slightly more pronounced than when using all Cα atoms (Table II). Based on the analysis of net-9 and net-6, almost half of the simulations reached a value smaller than their initial values although the average values are again very similar to the initial values. Since net-6 differs from net-9 by three residues located in β1 and β5, the differences between net-6 and net-9 RMSD reflect the degree to which the structure of those two strands is native-like.

Analysis of the local RMSD from the native state

To examine the extent to which native-like structures formed locally (independently from the global fold), RMSD values were determined for subsets of Cα atoms. The RMSD was calculated for each subset: (a) after a least square fit on the entire set of Cα atoms (all-Cα, global lsq fit); (b) after fitting only the subset of Cα atoms corresponding to the specific region of interest (local lsq fit). The former quantity reflects both the deformation of the structure of the subset itself and the

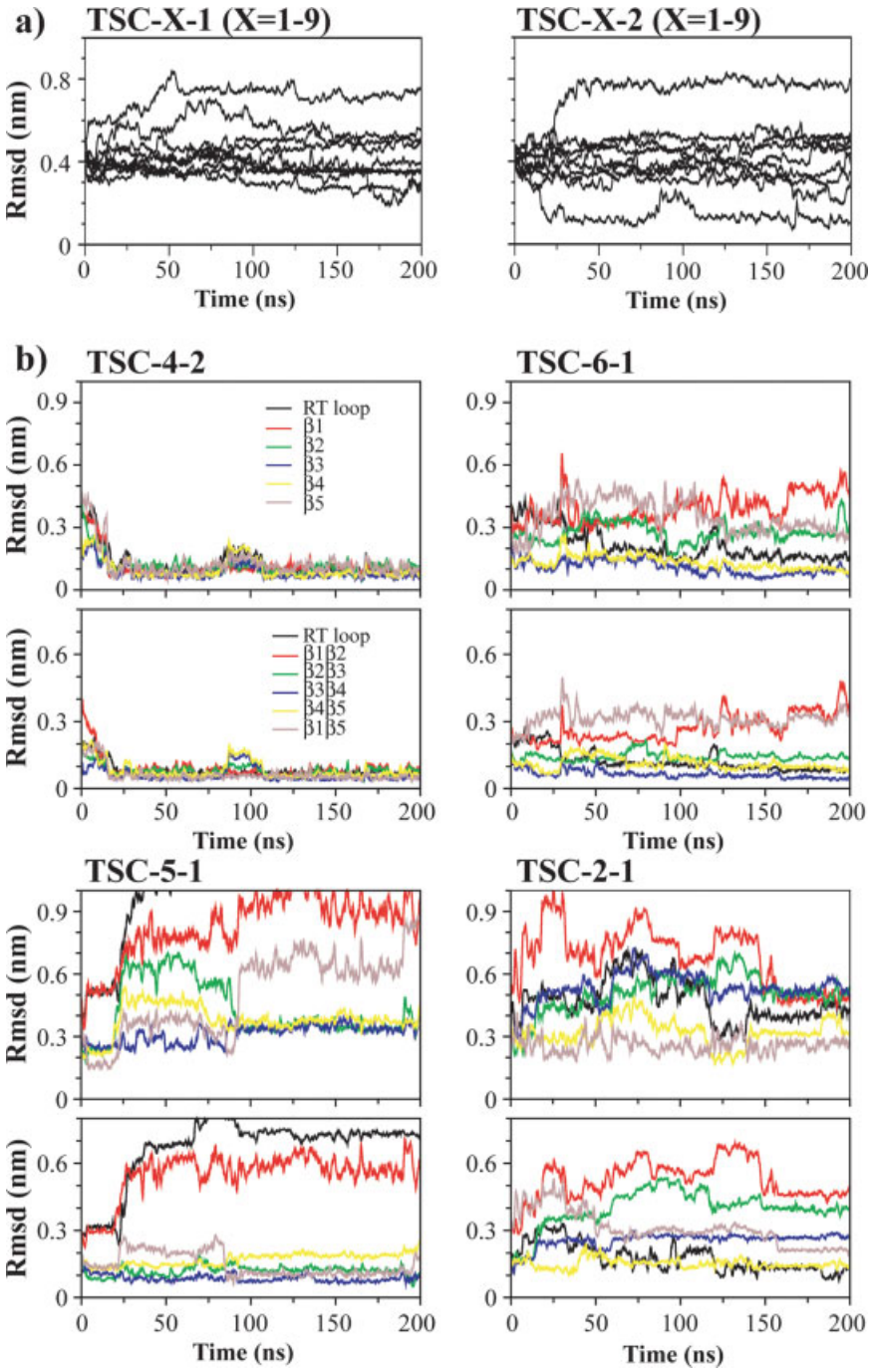


Figure 4

*RMSDs time series representing the molecular dynamics trajectories started from representative TSCs. (**a**) Global RMSD: root mean square deviation of all-Cα from the X-ray structure of the TSC-X-1 (left) and TSC-X-2 (right) with X = 1 to 9. (**b**) Local RMSD. A selection of four refolding scenarios is presented: complete folding (TSC-4-2), partial folding (TSC-6-1), unfolding with excess of secondary structure formation (TSC-5-2), and unfolding with little secondary structure formed (TSC-2-1). RMSD values of subsets of Cα atoms (see Method section for definition) after a least square fit on all-Cα subset (top panel) and on the same subset of Cα (bottom panel). The color code is as indicated in TSC-4-2. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]*

DOI 10.1002/prot

Table II
Local RMSD

TSC no.		1	2	3	4	5	6	7	8	9	<>	Native
All Cα	Init	4.3	4.5	4.1	4.4	4.2	4.8	4.1	5.1	5.1	4.5	1.4
	Run 1, 2	3.5 4.5	5.2 5.2	5.5 3.5	2.7 **1.1**	7.4 7.8	2.7 4.6	3.5 5.2	4.0 3.5	5.0 4.0	4.4	
Net 6	Init	1.7	2.3	2.5	1.6	1.8	2.2	2.0	2.7	2.9	2.2	0.7
	Run 1, 2	2.0 **1.1**	3.6 **1.5**	4.2 **0.9**	**1.5** **0.8**	4.7 3.0	**1.4** 1.6	1.7 3.5	2.5 **1.1**	4.0 1.9	2.3	
Net 9	Init	3.6	3.5	3.4	3.3	3.1	3.4	2.7	3.2	3.8	3.3	0.7
	Run 1, 2	2.7 2.9	3.3 4.3	3.9 2.2	1.7 **0.8**	5.4 4.2	1.8 2.4	3.1 4.2	2.5 2.7	3.9 1.8	3.0	
RT loop	Init	1.9	2.0	2.4	2.0	2.1	2.4	2.9	3.8	4.1	2.6	0.7
	Run 1, 2	2.0 2.9	**1.5** 3.9	4.3 1.6	1.9 **0.8**	5.0 7.1	**0.9** 2.8	2.6 3.8	4.5 2.7	4.7 4.1	3.2	
β1β2	Init	4.5	3.6	2.9	3.4	3.7	3.5	3.1	3.0	2.1	3.3	0.9
	Run 1, 2	3.0 3.1	5.2 3.9	4.4 1.9	**0.7** **0.8**	4.5 5.5	3.9 3.0	3.7 5.0	2.1 2.1	2.2 2.9	3.2	
β2β3	Init	1.9	2.1	2.2	1.6	2.0	2.3	1.6	1.6	1.3	1.9	0.8
	Run 1, 2	2.2 **1.3**	4.3 2.1	4.7 **0.8**	**0.6** **0.6**	3.1 **1.2**	**1.3** **0.8**	1.9 3.4	2.2 **1.3**	4.4 2.3	2.1	
β3β4	Init	1.3	1.2	1.7	1.4	1.2	1.5	1.5	1.3	1.3	**1.4**	0.5
	Run 1, 2	1.8 **0.8**	2.6 **0.7**	**0.9** **0.5**	**0.7** **0.5**	**1.5** **1.0**	**0.4** **0.7**	**0.8** **1.3**	**1.1** **0.7**	**0.7** **0.8**	**1.0**	
β4β5	Init	1.6	2.0	2.0	2.5	2.1	2.4	1.8	1.3	1.6	1.9	0.5
	Run 1, 2	1.6 3.7	**1.5** 1.9	**1.3** 1.7	**1.0** **0.8**	4.0 2.5	**1.0** 1.8	**1.1** 1.7	**1.4** 1.7	**1.0** **0.9**	1.7	
β1β5	Init	3.2	4.0	2.7	2.4	2.1	3.1	2.0	3.3	3.7	2.9	0.5
	Run 1, 2	2.5 3.2	2.7 4.8	1.7 4.0	**0.5** **0.6**	5.9 1.8	3.3 5.0	2.8 3.6	3.1 3.6	3.6 1.8	3.0	

Local root mean square deviations for subsets of Cα atoms (see Methods for definition). The values are for the conformations at the end of the 200 ns and are in Å. Bold characters indicate the subsets considered folded (rmsd $\leq$ 1.5 Å).

contribution due to changes in orientation in respect to the native structure. The latter quantity is a more direct measure of the degree to which the local structure is native like.

The local RMSD was calculated as a function of time for the four selected example cases [TSC-4-2, TSC-6-1, TSC-5-2, and TSC-2-1; Fig. 4(b)], which highlight how the degree to which native-like structure is retained on a local level can vary throughout the simulations. From Figure 4(b) it can also be seen that the folding of TSC-4-2 is highly cooperative involving all the subsets. Cooperative formation of native structure was also evident in TSC4-1 (results not shown). In the case of TSC-6-1, the RT loop, β3 and β4 folded within 60 ns, β2 and β5 stabilized with RSMD values around 3 Å, and β1 remained in a nonnative conformation. The local least-square fit plot suggests that the RT loop, β2β3, β3β4, and β4β5 locally folded with RMSD values <1.5 Å. The interactions formed by β1β2 and β1β5 were not native-like as is evident in the cartoon representation of the final conformation from TSC-6-1 (Fig. 3). TSC-5-2 shows the local folding of major parts of the protein (β2β3, β3β4, β4β5, and β1β5). However, based on its RMSD, this TSC is amongst the most unfolded with an all-Cα RMSD of 7.8 Å. In contrast, TSC-2-1 had a lower all-Cα RMSD but almost no native like local structure with only the RT loop relatively well folded.

The RMSD values for the different subsets of Cα after 200 ns of simulation are summarized for all TSCs in Table II and show how in the case of TSC-4-1, TSC-6-1, and TSC-3-2 the native fold was approached in the presence of nonnative conformations for the RT loop, β1, and β5, respectively. By contrast, TSC-5-1 can be classified as unfolded by almost all measures. In most of the remaining simulations only small regions of the molecule were observed to be native-like and the majority of the molecule remained unstructured, with the exception of the β3-β4 region, which became significantly more native-like in most cases.

RMSD per residue

A residue-level representation of the deviation of the structure with respect to the native structure as a function of time is presented in Figure 5, where the RMSD for each residue is shown after a least square fit on all-Cα; time runs from the back of the plot to the front, and the residue number is indicated on the front axes and the color code is related the RMSD value (*z*-axes). The RMSD from the X-ray structure for the native fold plus four selected cases (TSC-4-2, TSC-6-1, TSC-5-2, and TSC-2-1) is shown. In Figure 5 the dynamics and cooperative behavior of the folding/unfolding regions of the protein can be readily observed. Folding appeared to be cooperative for TSC-4-2 with almost the whole protein simultaneously reaching a RMSD value <1.5 Å at the initial stages (<20 ns) of the simulation. Local deformations occurred subsequently after about 100 ns, which involved residues 14–22 (i.e. part of the RT loop), and they are clearly visible as peaks in Figure 5. This figure illustrates also the flexibility of regions connecting elements of secondary structure, such as those extending from residue 24 to 25, 31 to 35, 40 to 45, and 51 to 52. In the case of TSC-6-1 the large deviations in β1, the diverging turn region (residues 21–23) and β5, are evident as are the

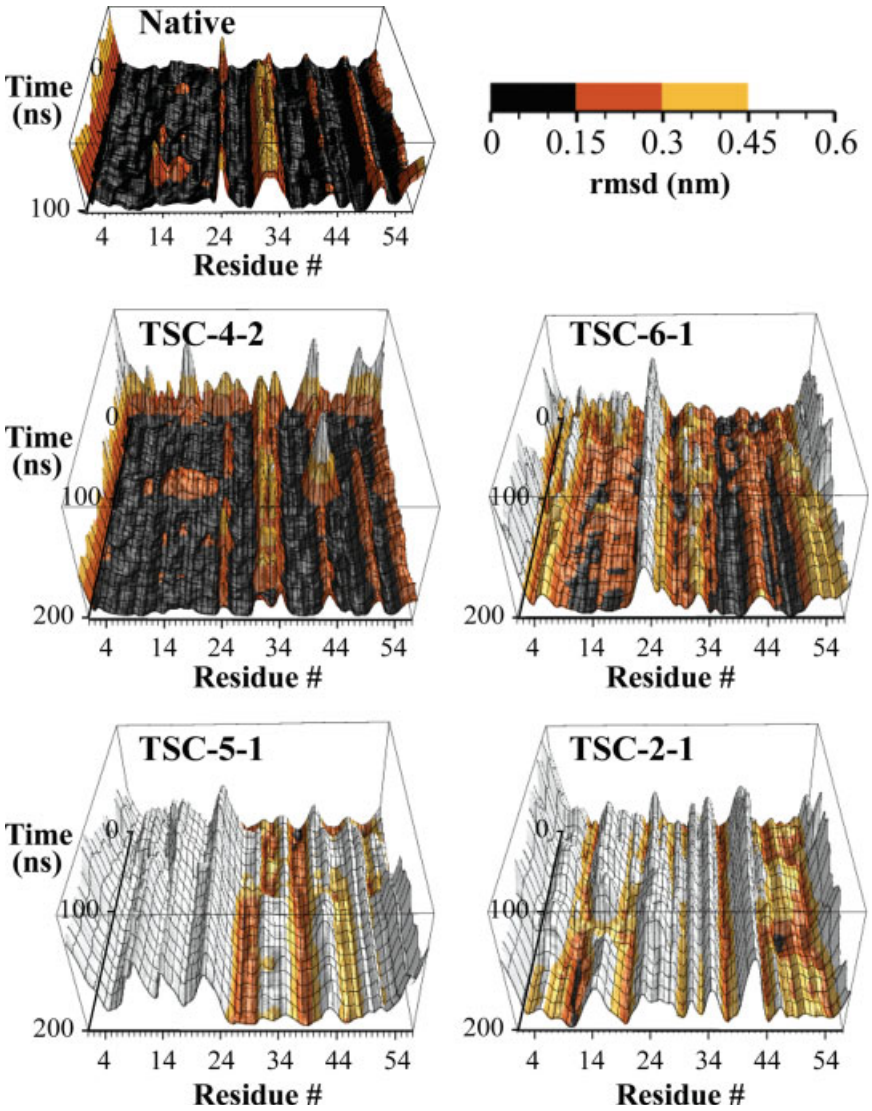


Figure 5

Time series of the RMSD plotted per residue. The native fold simulations and a selection of four refolding scenarios are presented: complete folding (TSC-4-2), partial folding (TSC-6-1), unfolding with excess of secondary structure formation (TSC-5-2), and unfolding with almost no secondary structure formed (TSC-2-1).

fully folded regions notably within the RT loop, β3 and β4. The plots of TSC-5-2 and TSC-2-1 clearly indicate that the protein has undergone global unfolding, although local arrangements were native-like. In these cases the RMSD per residue also contains a large contribution due to the least square fitting procedure.

Φ-Values

Φ-Values were estimated as the ratio of native contacts per residue, using a similar approach as that used by Lindorff-Larsen *et al.*[27] to generate the initial structures (see Methods). The Φ-values estimated in this way provide a measure of the degree to which the microenvironment of each residue is native-like. The averages of the Φ-values for TSC-X-1 and TSC-X-2 (Table III) indicate that there is a global increase of the average number of native contacts, from 53 to 61%, during the first 10 ns of the simulations. The average at the end of the simulations (last 10 ns) is similar, 65%, but the range of Φ-values observed at the end of the period simulated (0.34–0.94) is much wider than at the beginning of the simulation (0.44–

DOI 10.1002/prot

Table III
Φ-Values

TSC no.		1		2		3		4		5		6		7		8		9		< >
Φ-Value	Init	0.58		0.47		0.55		0.57		0.48		0.50		0.58		0.52		0.51		0.53
	0–10 ns	0.69	0.61	0.59	0.44	0.66	0.69	0.65	0.70	0.55	0.58	0.69	0.57	0.61	0.56	0.59	0.61	0.60	0.63	0.61
	190–200 ns	0.64	0.60	0.63	0.53	0.60	**0.83**	**0.84**	**0.94**	0.34	0.50	**0.77**	0.69	0.62	0.55	0.68	0.67	0.56	0.63	0.65

Φ-Values in the initial structure and averaged over the first and last 10 ns of the TSC-*X*-1 and TSC-*X*-2 simulations. < > indicates the average over the nine conformations. Bold characters indicate the values for folded and partially folded structures.

0.70). As expected, the TSC identified as folded or partially folded have the highest average Φ-values at the end of the simulation.

The initial increase of the average Φ-values is primarily due to changes in β1, β2, and the diverging turn and to a smaller extent in the region β2-4 and the RT loop, as evident in Figure 6(a–c). The β2-4 region has a native-like environment as shown by the average Φ-values for the last 10 ns of the simulations. In contrast, in the first half of the sequence the Φ-values are evenly dispersed in both sets of simulations with Φ-values ranging from 0.0 to 1.0 (averaging around 0.5).

The time series of Φ-values per residue for the native fold and for the four selected cases is presented in Figure 7. In the simulation of the native fold more than 96% of the native contacts are present throughout the simulation. In the region from residue 31–33 (n-src loop) Φ-values ranged from 0.4 to 0.6, indicating a loss of native contacts. This finding is consistent with the observation that the n-src loop region is flexible.[39] A similar behavior was observed for TSC-4-2 after folding. For TSC-6-1, native contacts rapidly formed within the RT loop and the initial high ratio of native contacts was conserved throughout the simulation. By contrast, nonnative conformations were observed in β1 and in the diverging turn (residues 20–24). In TSC-5-2, few native contacts were present in the first half of the sequence (residues 2–30). In the remaining sequence (mainly β2–β4) many native like contacts have formed. In contrast, in TSC-2-1, native contacts are distributed along the sequence. Overall, the observations based on the Φ-values are consistent with the RMSD analysis discussed earlier. However, in some cases, high Φ-values do not indicate a low RMSD, and vice versa, as illustrated in the cases of TSC-6-1 and TSC-2-1 [Figs. 4(b), 5, 6, and Table II], suggesting that RMSD and Φ-values provide complementary information.

Statistics of folding for two TSCs

According to the results shown in Table II one would suggest that TSC-4 is on the folded side of the transition state and TSC-5 on the unfolded side, while no trend is apparent for the other structures. However, these suggestions are based only on two simulations. To further assess their statistical significance, eight additional 200 ns simulations were performed for TSC-4 and TSC-5 giving a total of 10 independent simulations for each of these two cases. TSC-4 and TSC-5 were chosen as they showed the greatest difference in folding behavior (see Tables I–III).

The results for the 10 independent 200 ns simulations of TSC-4 and TSC-5 are summarized in Table IV, which shows two notable features. The first is the wide variation between the results in the different runs. The second is that despite these wide variations there are systematic differences between TSC-4 and TSC-5. Although the RMSD value of the starting structure in the case of TSC-4 is slightly larger than that of TSC-5 (4.4–4.2 Å), on average the TSC-4 moves toward the native state by about 1.0 Å, while TSC-5 moves away from the native state by about 1.0 Å. In both TSC-4 and TSC-5, the region of β3 and β4 shows a clear propensity to form a native-like structure. With the possible exception of the RT loop in TSC-5, for which the RMSD value is significantly larger than the initial structure in all 10 simulations, none of the local elements showed the same behavior in all simulations. It is also notable that the range of values observed for both TSC-4 and TSC-5 overlap, and that they are similar to the range of values observed for the other TSCs.

Criteria for the identification of folded and unfolded states

The determination of the folding propensity of a TSC requires a criterion for discriminating between folded and unfolded structures. By considering cases such as TSC-4-2, for which complete folding was observed, the discrimination would appear to be straightforward. This case also suggests that if the initial structure is on the folding pathway, a simulation of the folding from a structure with an initial all-Cα RMSD >4.0 Å can be achieved using appropriate atomic models and explicit solvent. TSC-4-2 also highlights the fact that a decrease in all-Cα RMSD to values of 3.0 Å or even 2.0 Å from the experimentally determined structure in this case does not necessarily correspond to a fully folded state. In the current case, sampling of states with an all-Cα RMSD of <1.5 Å with at least 90% of native contacts formed would be required to ensure the protein had in fact fully folded.

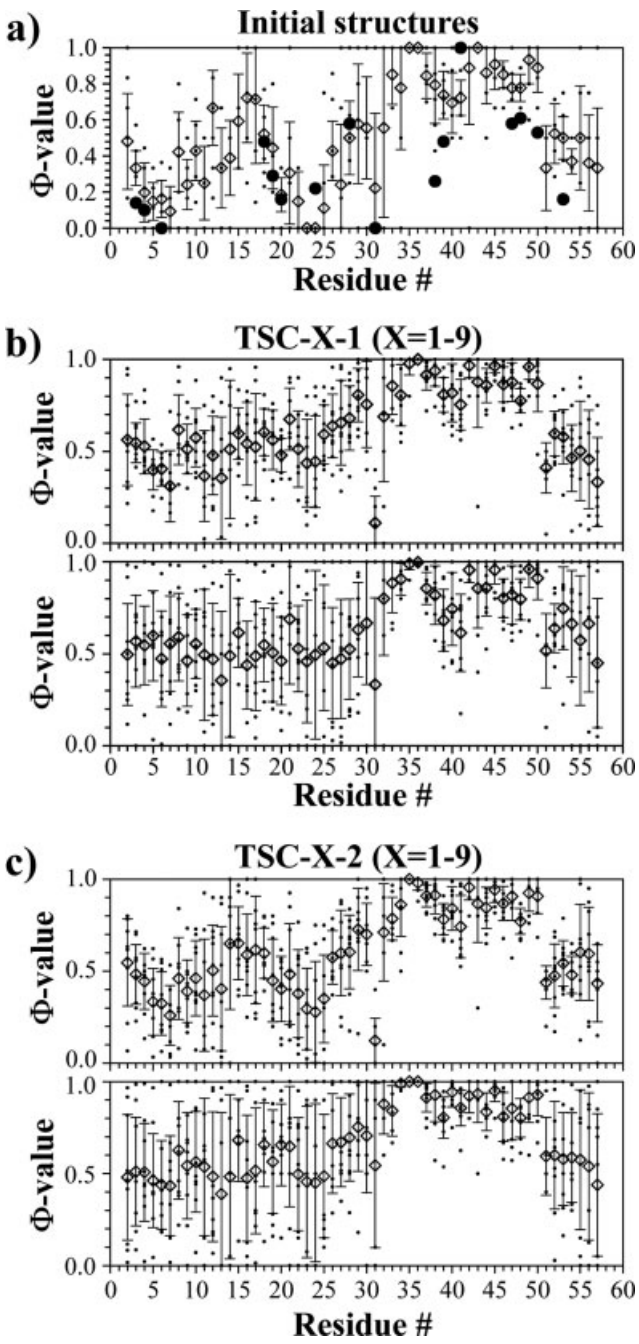


Figure 6

*Calculated Φ-values (see Methods) for (**a**) the initial structures, (**b**) for the TSC-X-1, (**c**) for the TSC-X-2. X = 1 to 9. In panel (**b**) and (**c**) the top panel is the average over the first 10 ns, and the bottom panel over the last 10 ns. In all panels each individual value is indicated by a dot and the average is given (diamond) with one standard deviation. The experimental Φ-values used in the construction of the TSE[27] are shown in panel (**a**) by big black dots.*

The definition of an unfolded state is in many respects even more problematic than that of the folded state. This is because the unfolded state, despite recent advances,[40–44] cannot be currently characterized experimentally with the same structural detail that is possible for native states, and as a result it is more subjective to define criteria for classifying a particular conformation as unfolded. For example, in the case of the drkN SH3 domain,[43] it has been suggested based on SAXS and NMR data that a 30–40% increase of R_{gyr} could be expected in the unfolded state. By adopting this criterion, none of the TSC would be considered as unfolded during the 200 ns presented here. Similarly, based on the values of Cα RMSD, R_{gyr}, and SASA for post and pre-TS conformations extracted from a previous Pfold analysis of Shakhnovich and coworkers using a Go model,[20] all of the structures simulated would again be on the folding side of the TS. Gsponer and Caflisch[13] used an all-Cα RMSD of <2.5 or >7 Å and >87.5 or <37.5% of native contacts as structural criteria to differentiate between folding and unfolding simulations for a putative TSC for the src SH3 domain. By these definitions, in the present case only TSC-4-1 would be classified as being folded and only TSC-5-1 as unfolded, while the remaining structures would remain undetermined.

Taken together these observations suggest that most of the TSCs showed comparable tendencies to either fold or unfold. In this respect, the TSCs analyzed can be considered putative members of the TSE. In the following section we describe why a Pfold analysis was not possible in the present case.

Possible effects caused by the presence of folding intermediates

The nine transition state configurations analyzed in this study have similar structural characteristics, in terms of topology, radius of gyration, and solvent accessible surface area. However, even in the case of TSC-4, for which complete folding events were observed in our simulations, a wide variety of structures were obtained in replicate simulations that, although very different from the starting structure, were neither folded nor unfolded. The inability to reach the native state from the transition state within 200 ns may be a consequence of the intermediate states that have been recently detected using relaxation dispersion NMR techniques for mutational variants of the Fyn and the Abp1p SH3 domains.[45,46] In both cases, the conversion from the intermediates to the native state was observed to take place on the millisecond timescale. Therefore, folding from the transition state in the case of SH3 proteins appears to take place on a landscape characterized by intermediate states whose lifetimes may span timescales from ns to ms. Based on the observation that only one of the 34 simulations performed reached the native state within 200 ns, first-order reaction rate theory suggests that on average trajectories starting from the transition state would require tens of microseconds to reach the native state. The presence of metastable states along the folding pathway should then be

DOI 10.1002/prot

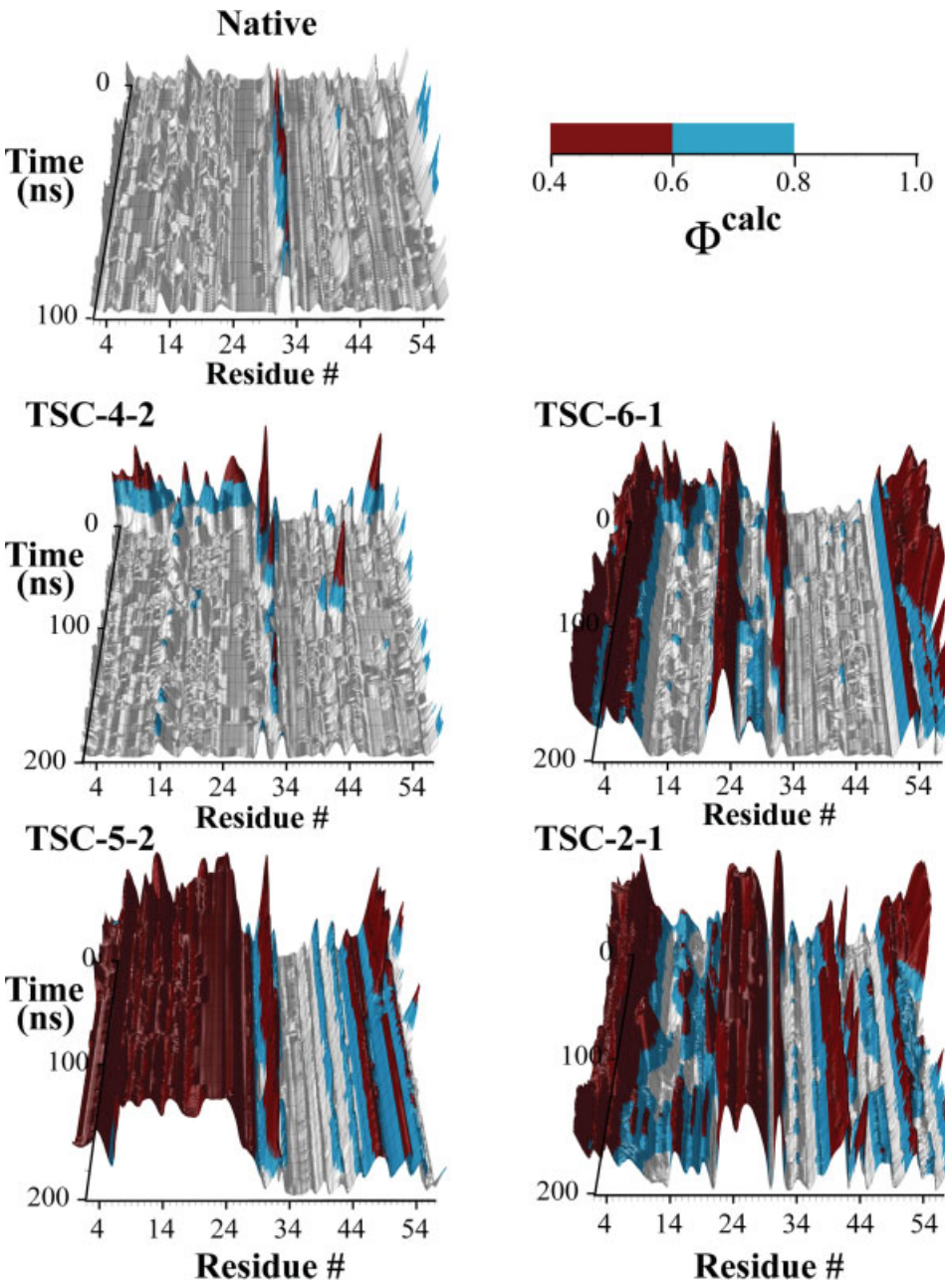


Figure 7

Time series of calculated Φ-values per residue for the four selected cases: complete folding (TSC-4-2), partial folding (TSC-6-1), unfolding with excess of secondary structure formation (TSC-5-2), and unfolding with little secondary structure formed (TSC-2-1). [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

accounted for when using of the Pfold approach to validate structural models for transition states.

Effects that require further studies

We mention three factors that may influence the conclusions that we propose in the current study and that will require additional studies. (1) The TSCs that we considered were generated by a procedure in which the assumption was made that the experimental Φ-values correlate with the ratio of native contacts in the TSE.[8–13] For individual Φ-values this assumption may not be correct, and recent studies have suggested that in some cases the contacts contributing to the Φ-values may be nonnative.[25,47] Approaches such as those in which free

Table IV
Local RMSD (in Å) for the TSC-4-Y and TSC-5-Y Simulations

		TSC-4-*Y*											TSC-5-*Y*											
Y =		1	2	3	4	5	6	7	8	9	10	< >	1	2	3	4	5	6	7	8	9	10	< >	Native
All Cα	Init					**4.4**											**4.2**							**1.4**
		2.7	1.1	5.3	2.7	2.8	3.9	3.6	4.8	3.1	4.5	**3.4**	7.4	7.8	3.8	3.8	4.4	5.7	4.3	5.6	5.3	3.9	**5.2**	
Net 6	Init					**1.6**											**1.8**							**0.7**
		1.5	0.8	1.9	2.1	1.5	2.9	1.4	3.7	1.9	2.9	**2.1**	4.7	3.0	3.0	2.7	2.0	3.2	1.5	2.4	2.2	1.8	**2.6**	
Net 9	Init					**3.3**											**3.1**							**0.7**
		1.7	0.8	3.1	2.2	1.8	3.0	1.9	4.2	2.2	3.5	**2.4**	5.4	4.2	2.6	2.7	2.9	5.3	1.5	4.2	4.2	2.0	**3.5**	
RT loop	Init					**2.0**											**2.1**							**0.7**
		1.9	0.8	2.1	1.4	2.6	1.0	1.6	2.8	2.8	1.5	**1.8**	5.0	7.1	4.6	3.2	3.7	2.9	3.3	4.1	3.0	2.9	**4.0**	
β1β2	Init					**3.4**											**3.7**							**0.9**
		0.7	0.8	3.3	2.5	2.1	3.6	1.5	4.2	2.6	3.8	**2.5**	4.5	5.5	3.2	3.1	2.7	7.5	2.4	3.5	3.6	2.5	**3.8**	
β2β3	Init					**1.6**											**2.0**							**0.8**
		0.6	0.6	2.1	1.8	1.3	4.1	1.3	3.1	1.8	3.1	**2.0**	3.1	1.2	2.9	2.7	1.1	1.2	2.8	1.0	1.7	2.3	**2.0**	
β3β4	Init					**1.4**											**1.2**							**0.5**
		0.7	0.5	1.9	0.6	0.5	0.5	0.9	0.7	0.6	1.2	**0.8**	1.5	1.0	1.2	0.8	0.9	0.8	0.9	0.8	1.6	0.8	**1.0**	
β4β5	Init					**2.5**											**2.1**							**0.5**
		1.0	0.8	2.9	0.9	0.7	1.9	1.3	1.3	0.9	2.0	**1.4**	4.0	2.5	0.8	1.3	1.0	1.9	1.0	3.6	2.4	0.9	**1.9**	
β1β5	Init					**2.4**											**2.1**							**0.5**
		0.5	0.6	4.4	1.8	2.0	2.9	1.7	2.4	1.0	3.0	**2.0**	5.9	1.8	0.5	2.1	2.2	3.9	0.7	5.2	4.9	2.9	**3.0**	

Init indicates the value for the TSC. < > indicates the average over the 10 simulations. The Native column reports the values averaged over the 100 ns simulation of the native fold.

energy[48,49] or kinetic[25] evaluations of Φ-values are used will increase our understanding of these situations. (2) The TSE analyzed was generated at high temperature, using an implicit model of the solvent and in simulations in which the Φ-values restraints were all enforced simultaneously.[27] This procedure might introduce strain in the TSE structures, to which MD simulations are very sensitive. The use of ensemble-averaged Φ-value restraints[21,50,51] and simulated annealing cycles[17] have been proposed to reduce these effects. (3) The force field used, in this case GROMOS 41a3 in conjunction with the SPC water model, should be capable of accurately describing of the conformational free energy landscape of the protein, as suggested by the fact that we observed the complete folding of one TSC from a 4.4 Å to about 1.1 Å RMSD. Nevertheless, the reliability of any force field in describing the relative statistical weights of the unfolded, intermediate and native states, but also the energetic barriers between them is open to question, and especially relevant to the present case in which intermediate states appear to play an important role.

DISCUSSION AND CONCLUSIONS

Nine conformations selected from a model of the TSE for the α-spectrin SH3 domain[27] have been simulated in explicit solvent in order to analyze their refolding behaviour. In all cases a clear regularization of SSEs was observed, with local regions of native-like structure appearing in most cases. This ordering process was concomitant with a decrease of both the radius of gyration and the solvent exposed surface area of the protein. In particular, two elements of secondary structure, β3 and β4, showed a strong preference to acquire their native conformation. By contrast, three other elements, β1, β2, and β5, showed a similar tendency to either fold or unfold, whereas on average the RT loop was observed to unfold during the simulations. In one case (TSC-4-2) we detected a folding event with the all-Cα RMSD from the native structure decreasing from 4.4 Å to about 1.1 Å; this process occurred within 20 ns and was highly cooperative. Once the native fold was reached the structure remained stable for the rest of the simulation (180 ns); partial cooperative folding was observed in several other cases. These results suggest that the native fold of α-spectrin SH3 domain corresponds to a free energy minimum in the GROMOS 43a1 force field, which therefore appear to be capable of predicting the folding of this protein.

The results that we have presented highlight the difficulty of unambiguously distinguishing between folding and unfolding trajectories. We have shown that in this case, even seemingly rather conservative criteria, such as an all-Cα RMSD of <2.5 Å and a fraction >85% of native contacts,[13] would not guarantee the achievement of a fully folded structure; to ensure this result, values such as an all-Cα RMSD of <1.5 Å and a fraction >90% of native contacts are required.

In summary, this study of α-spectrin SH3 domain has shown that even starting from structures with a native topology representing a model of the TSE for folding of this protein, 200 ns of simulation are not sufficient to observe folding or unfolding in a consistent manner.

DOI 10.1002/prot

Therefore, our results suggest the presence in the case of SH3 domain proteins of intermediates on a submillisecond timescale along the folding pathway in addition to those on the millisecond timescale already identified experimentally.[45,46]

ACKNOWLEDGMENTS

The authors thank K. Lindorff-Larsen for providing the structures of the transition state of the α-spectrin SH3 domain and E. Paci for useful comments.

REFERENCES

1. Anfinsen CB, Haber E, Sela M, White FH. Kinetics of formation of native ribonuclease during oxidation of reduced polypeptide chain. Proc Natl Acad Sci USA 1961;47:1309–1314.
2. Dinner AR, Sali A, Smith LJ, Dobson CM, Karplus M. Understanding protein folding via free-energy surfaces from theory and experiment. Trends Biochem Sci 2000;25:331–339.
3. Dill KA, Chan HS. From Levinthal to pathways to funnels. Nat Struct Biol 1997;4:10–19.
4. Onuchic JN, Wolynes PG. Theory of protein folding. Curr Opin Struct Biol 2004;14:70–75.
5. Fersht AR, Daggett V. Protein folding and unfolding at atomic resolution. Cell 2002;108:573–582.
6. Fersht AR, Leatherbarrow RJ, Wells TNC. Quantitative-analysis of structure-activity-relationships in engineered proteins by linear free-energy relationships. Nature 1986;322:284–286.
7. Fersht AR. Structure and mechanism in protein science: a guide to enzyme catalysis and protein folding. New York: W. H. Freeman; 1999.
8. Li AJ, Daggett V. Characterization of the transition-state of protein unfolding by use of molecular-dynamics—chymotrypsin inhibitor-2. Proc Natl Acad Sci USA 1994;91:10430–10434.
9. Clementi C, Nymeyer H, Onuchic JN. Topological and energetic factors: What determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? An investigation for small globular proteins. J Mol Biol 2000; 298:937–953.
10. Vendruscolo M, Paci E, Dobson CM, Karplus M. Three key residues form a critical contact network in a protein folding transition state. Nature 2001;409:641–645.
11. Li L, Shakhnovich EI. Constructing, verifying, and dissecting the folding transition state of chymotrypsin inhibitor 2 with all-atom simulations. Proc Natl Acad Sci U S A 2001;98:13014–13018.
12. Paci E, Vendruscolo M, Dobson CM, Karplus M. Determination of a transition state at atomic resolution from protein engineering data. J Mol Biol 2002;324:151–163.
13. Gsponer J, Caflisch A. Molecular dynamics simulations of protein folding from the transition state. Proc Natl Acad Sci USA 2002;99:6719–6724.
14. Settanni G, Gsponer J, Caflisch A. Formation of the folding nucleus of an SH3 domain investigated by loosely coupled molecular dynamics simulations. Biophys J 2004;86:1691–1701.
15. Paci E, Clarke J, Steward A, Vendruscolo M, Karplus M. Self-consistent determination of the transition state for protein folding: Application to a fibronectin type III domain. Proc Natl Acad Sci USA 2003;100:394–399.
16. Paci E, Lindorff-Larsen K, Dobson CM, Karplus M, Vendruscolo M. Transition state contact orders correlate with protein folding rates. J Mol Biol 2005;352:495–500.
17. Salvatella X, Dobson CM, Fersht AR, Vendruscolo M. Determination of the folding transition states of barnase by using Phi(I)-value-restrained simulations validated by double mutant Phi(IJ)-values. Proc Natl Acad Sci USA 2005;102:12389–12394.
18. Du R, Pande VS, Grosberg AY, Tanaka T, Shakhnovich EI. On the transition coordinate for protein folding. J Chem Phys 1998;108: 334–350.
19. Ding F, Guo WH, Dokholyan NV, Shakhnovich EI, Shea JE. Reconstruction of the src-SH3 protein domain transition state ensemble using multiscale molecular dynamics simulations. J Mol Biol 2005; 350:1035–1050.
20. Hubner IA, Edmonds KA, Shakhnovich EI. Nucleation and the transition state of the SH3 domain. J Mol Biol 2005;349:424–434.
21. Latzer J, Eastwood MP, Wolynes PG. Simulation studies of the fidelity of biomolecular structure ensemble recreation. J Chem Phys 2006;125:214905–214916.
22. Cavalli A, Vendruscolo M, Paci E. Comparison of sequence-based and structure-based energy functions for the reversible folding of a peptide. Biophys J 2005;88:3158–3166.
23. Rhee YM, Pande VS. On the role of chemical detail in simulating protein folding kinetics. Chem Phys 2006;323:66–77.
24. Rao F, Settanni G, Guarnera E, Caflisch A. Estimation of protein folding probability from equilibrium simulations. J Chem Phys 2005;122:184901–184905.
25. Settanni G, Rao F, Caflisch A. phi-Value analysis by molecular dynamics simulations of reversible folding. Proc Natl Acad Sci USA 2005;102:628–633.
26. Lenz P, Zagrovic B, Shapiro J, Pande VS. Folding probabilities: A novel approach to folding transitions and the two-dimensional Ising-model. J Chem Phys 2004;120:6769–6778.
27. Lindorff-Larsen K, Vendruscolo M, Paci E, Dobson CM. Transition states for protein folding have native topologies despite high structural variability. Nat Struct Mol Biol 2004;11:443–449.
28. Martinez JC, Pisabarro MT, Serrano L. Obligatory steps in protein folding and the conformational diversity of the transition state. Nat Struct Biol 1998;5:721–729.
29. Lindahl E, Hess B, van der Spoel D. GROMACS 3.0: a package for molecular simulation and trajectory analysis. J Mol Model 2001;7: 306–317.
30. Berendsen HJC, van der Spoel D, van Drunen R Gromacs—a message-passing parallel molecular-dynamics implementation. Comput Phys Commun 1995;91:43–56.
31. van Gunsteren WF, Billeter SR, Eising AA, Hunenberger PH, Kruger P, Mark AE, Scott WRP, Tironi IG. Biomolecular simulation: The GROMOS96 manual and user guide. Vdf Hochschulverlag AG an der ETH Zurich and BIOMOS b.v., Zurich, Groningen, 1996.
32. Berendsen HJC, Postma JPM, van Gunsteren WF, Hermanns J. Interaction models for water in relation to protein hydration. In: Pullman B, editor. Intermolecular forces. Dordrecht: Reidel; 1981. pp 331–342.
33. Berendsen HJC, Postma JPM, van Gunsteren WF, Dinola A, Haak JR. Molecular-dynamics with coupling to an external bath. J Chem Phys 1984;81:3684–3690.
34. Tironi IG, Sperb R, Smith PE, van Gunsteren WF. A Generalized reaction field method for molecular-dynamics simulations. J Chem Phys 1995;102:5451–5459.
35. Hess B, Bekker H, Berendsen HJC, Fraaije J. LINCS: A linear constraint solver for molecular simulations. J Comput Chem 1997;18: 1463–1472.
36. Miyamoto S, Kollman PA. Settle—an analytical version of the shake and rattle algorithm for rigid water models. J Comput Chem 1992; 13:952–962.
37. Eisenhaber F, Lijnzaad P, Argos P, Sander C, Scharf M. The Double cubic lattice method—efficient approaches to numerical-integration of surface-area and volume and to dot surface contouring of molecular assemblies. J Comput Chem 1995;16:273–284.
38. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 1983;22:2577–2637.
39. Sadqi M, Casares S, Abril MA, Lopez-Mayorga O, Conejero-Lara F, Freire E. The native state conformational ensemble of the SH3 domain from α-spectrin. Biochemistry 1999;38:8899–8906.

40. Klein-Seetharaman J, Oikawa M, Grimshaw SB, Wirmer J, Duchardt E, Ueda T, Imoto T, Smith LJ, Dobson CM, Schwalbe H. Long-range interactions within a nonnative protein. Science 2002;295: 1719–1722.
41. Religa TL, Markson JS, Mayor U, Freund SMV, Fersht AR. Solution structure of a protein denatured state and folding intermediate. Nature 2005;437:1053–1056.
42. Cho JH, Raleigh DP. Electrostatic interactions in the denatured state and in the transition state for protein folding: Effects of denatured state interactions on the analysis of transition state structure. J Mol Biol 2006;359:1437–1446.
43. Choy WY, Mulder FAA, Crowhurst KA, Muhandiram DR, Millett IS, Doniach S, Forman-Kay JD, Kay LE. Distribution of molecular size within an unfolded state ensemble using small-angle X-ray scattering and pulse field gradient NMR techniques. J Mol Biol 2002; 316:101–112.
44. Kohn JE, Millett IS, Jacob J, Zagrovic B, Dillon TM, Cingel N, Dothager RS, Seifert S, Thiyagarajan P, Sosnick TR, Hasan MZ, Pande VS, Ruczinski I, Doniach S, Plaxco KW. Random-coil behavior and the dimensions of chemically unfolded proteins. Proc Natl Acad Sci USA 2004;101:12491–12496.
45. Korzhnev DM, Salvatella X, Vendruscolo M, Di Nardo AA, Davidson AR, Dobson CM, Kay LE. Low-populated folding intermediates of Fyn SH3 characterized by relaxation dispersion NMR. Nature 2004;430:586–590.
46. Korzhnev DM, Neudecker P, Zarrine-Afsar A, Davidson AR, Kay LE. Abp1p and Fyn SH3 domain fold through similar low-populated intermediate sates. Biochemistry 2006;45:10175–10183.
47. Feng HQ, Vu ND, Zhou Z, Bai YW. Structural examination of Phi-value analysis in protein folding. Biochemistry 2004;43:14325–14331.
48. Lindorff-Larsen K, Paci E, Serrano L, Dobson CM, Vendruscolo M. Calculation of mutational free energy changes in transition states for protein folding. Biophys J 2003;85:1207–1214.
49. Gsponer J, Hopearuoho H, Whittaker SBM, Spence GR, Moore GR, Paci E, Radford SE, Vendruscolo M. Determination of an ensemble of structures representing the intermediate state of the bacterial immunity protein Im7. Proc Natl Acad Sci USA 2006;103:99–104.
50. Davis R, Dobson CM, Vendruscolo M. Determination of the structures of distinct transition state ensembles for a β-sheet peptide with parallel folding pathways. J Chem Phys 2002;117:9510–9517.
51. Best RB, Vendruscolo M. Structural interpretation of hydrogen exchange protection factors in proteins: characterization of the native state fluctuations of C12. Structure 2006;14:97–106.

 DOI 10.1002/prot