Formation of the Folding Nucleus of an SH3 Domain Investigated by Loosely Coupled Molecular Dynamics Simulations

G. Settanni, J. Gsponer, and A. Caflisch
Biochemisches Institut, Universität Zürich, Zürich, Switzerland

ABSTRACT The experimentally well-established folding mechanism of the *src*-SH3 domain, and in particular the ϕ-value analysis of its transition state, represents a sort of testing table for computational investigations of protein folding. Here, parallel molecular dynamics simulations of the *src*-SH3 domain have been performed starting from denatured conformations. By rescuing and restarting only trajectories approaching the folding transition state, an ensemble of conformations was obtained with a completely structured central β-sheet and a native-like packing of residues Ile-110, Ala-121, and Ile-132. An analysis of the trajectories shows that there are several pathways leading to the formation of the central β-sheet whereas its two hairpins form in a different but consistent way.

INTRODUCTION

Proteins can fold by simple two-state kinetics (Jackson and Fersht, 1991). Hence, if the native structure is known, an accurate description of the denatured state and transition state ensemble (TSE) should permit a complete understanding of two-state folding. According to the nucleation-condensation mechanism the TSE is characterized by an extensive network of interactions in regular secondary structure elements as well as a number of tertiary contacts (Daggett and Fersht, 2003). Unfortunately, it is not easy to determine and analyze the TSE with experimental techniques, since it corresponds to an unstable high-energy region on the free energy surface. One method based on protein engineering and pioneered by Fersht and co-workers provides structural information about the rate-limiting step of folding (Matouschek et al., 1989). Insights on the formation of side chain interactions in the TSE are obtained by deleting parts of individual residues and assessing the effect on folding kinetics and stability. The ϕ_i^{exp} value is defined as the ratio $\Delta\Delta G_{TS-D}/\Delta\Delta G_{N-D}$, where $\Delta\Delta G_{TS-D}$ is the change in free energy difference between the TSE and the denatured state induced by a mutation of residue i, and $\Delta\Delta G_{N-D}$ the change in free energy difference between the native state and the denatured state due to the same mutation. The ϕ_i^{exp} value is an indicator of the nativeness of residue i in the TSE: a ϕ_i^{exp} value of 1 indicates that residue i has a native-like structure in the TSE, whereas a value of 0 implies that, in the TSE, residue i is as unfolded as in the denatured state.

The TSE of chymotrypsin inhibitor 2 has been characterized by a synergistic interplay of experimental techniques (Φ-value analysis and NMR studies) and atomistic molecular dynamics (MD) simulations of protein unfolding at high temperature (Li and Daggett, 1996). The transition state structures identified in the simulations are consistent with the available experimental data and were used for interpreting ϕ_i^{exp} values at atomic level of detail (Daggett et al., 1996). The successful combination of experimental data and MD simulations allowed the description of the TSE of chymotrypsin inhibitor 2 as a state close to the native structure with an almost intact α-helix and a quite disrupted β-sheet. Furthermore, a folding nucleus consisting of Ala-16 in the α-helix and Leu-49 and Ile-57 in the β-sheet was identified (Daggett et al., 1996).

Monte Carlo simulations of lattice models have also been used to investigate the TSE (Dinner and Karplus, 1999a; Du et al., 1998; Li et al., 2000; Ozkan et al., 2001). Their efficiency allows the validation of putative transition state conformations by calculating the transmission coefficient, which is the probability of a given structure to fold before it unfolds. The transmission coefficient should be close to 0.5 for conformations in the TSE. The main drawback of lattice models is the coarse description of the protein structure and interactions. Off-lattice simulations of a C_α model with a Go potential have allowed the construction of the TSE of acylphosphatase from published ϕ_i^{exp} values (Vendruscolo et al., 2001). Furthermore, they have shed light on some aspects of the folding of SH3 (Borreguero et al., 2002; Ding et al., 2002), including the role of the desolvation process (Cheung et al., 2002); however, they do not allow the evaluation of the importance of non-native interactions.

The protein folding TSE arising from experiments and calculations performed on small two-state folding proteins is generally a distorted or expanded version of the native state (Daggett and Fersht, 2003; Schymkowitz et al., 2002), and little is known about the early phase of folding. Hence, the present study was motivated by the following question: is it possible to analyze the precritical phase of the folding process, i.e., from the denatured to the TSE? A positive answer to this question will also help in clarifying what the predominant driving forces are in protein folding. To address this issue, a possible approach is represented by atomistic MD simulation with an implicit representation of the solvent.

Submitted July 29, 2003, and accepted for publication October 22, 2003.

Address reprint requests to A. Caflisch, Biochemisches Institut, Universität Zürich, Winterthurerstrasse 190, 8057 Zürich, Switzerland. Fax: 411-635-6862; E-mail: caflisch@bioc.unizh.ch.

0006-3495/04/03/1691/11 $2.00

Recently, MD simulation technique has allowed a statistically relevant analysis of the reversible folding of structured peptides (Ferrara and Caflisch, 2000, 2001), of the unfolding of small proteins at moderate temperatures (Gsponer and Caflisch, 2001) even with explicit water molecules (Mayor et al., 2003), and of the folding of the *src*-SH3 domain from the TSE (Gsponer and Caflisch, 2002). The all-β-domain *src*-SH3 is a two-state folder with a native state structure consisting of a β-hairpin (formed by the terminal segments) packed orthogonally on top of a three-stranded antiparallel β-sheet. Here, starting from the denatured state of *src*-SH3, we run loosely coupled parallel simulations which are periodically stopped and restarted from the snapshot closest to the experimentally determined TSE. Following previous works by Li and Daggett (1996, 1994) and Vendruscolo et al. (2001), ϕ_i^{exp} values are interpreted in terms of native-like side-chain contacts. The main difference between the present approach and that of Li and Daggett (1996, 1994) is that the latter employs a conformational cluster analysis to identify TSE structures along unfolding trajectories whereas in the present study
we use experimental information to efficiently sample the transition from the denatured state to the TSE. The present approach differs from Vendruscolo et al. (2001) in several aspects. Here, instead of considering only the C_α positions, along with a Go-like force field and a Monte Carlo sampling, most of the atomic degrees of freedom are taken into account and a more accurate CHARMM empirical potential is used in the MD simulations. This allows us to extend the investigation to the pathways from the denatured state to the TSE. We anticipate that non-native contacts are observed using the present approach whereas they are penalized in the Go model. Furthermore, since the constraints defined by ϕ-values enclose a considerable region of the phase space of the protein, accurate energetics may allow the sampling of only the parts of this region that are more relevant, excluding those parts that present unphysically high energies. The main drawback of the present approach is that the required simulation time is larger than in the case of the C_α Go-model.

A large part of the secondary and tertiary structure of SH3, including the putative folding nucleus (Grantcharova et al., 2000; Northey et al., 2002a; Riddle et al., 1999), assumes the correct three-dimensional conformation during the simulations presented here. The formation of the folding nucleus of SH3 is reported with atomistic detail, and comparison with experimental data (mutational studies) is presented.

THEORY AND METHODS

MD simulations of *src*-SH3 domain

All the calculations were carried out by using an extended-atom model of the *src*-SH3 domain (Xu et al., 1997, PDB 1FMK) and the CHARMM program (Brooks et al., 1983). Solvation effects were approximated by an implicit model based on the solvent-accessible surface area (Ferrara et al., 2002). In this approximation, the solvation free energy is given by

$$G_{solv}(\mathbf{r}) = \sum_{i=1}^{N} \sigma_i A_i(\mathbf{r}), \quad (1)$$

for a molecule having N heavy atoms with Cartesian coordinates $\mathbf{r} = (\mathbf{r}_1, \ldots, \mathbf{r}_N)$. $A_i(\mathbf{r})$ is the solvent-accessible surface computed by an approximate analytical expression (Hasel et al., 1988) and using a 1.4 Å probe radius. Furthermore, ionic side chains were neutralized (Lazaridis and Karplus, 1999) and a linear distance-dependent screening function ($\varepsilon(r) = 2r$) was used for the electrostatic interactions. The CHARMM PARAM19 default cutoffs for long-range interactions were used, i.e., a shift function (Brooks et al., 1983) was employed with a cutoff at 7.5 Å for both the electrostatic and van der Waals terms. This cutoff length was chosen to be consistent with the parameterization of the force field. The model contains only two σ-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., 2002). The σ-parameters do not have a temperature-dependence, inasmuch as this was shown to be weak in the 330–360 K range by a previous implicit solvent model calibrated on amino acid hydration free energies (Elcock and McCammon, 1997). The model is not biased toward any particular secondary structure type. In fact, exactly the same force field, implicit solvation model, and values of the σ-parameters have been used in MD simulations of folding of structured peptides (α-helices and β-sheets) ranging in size from 15 to 31 residues (Ferrara and Caflisch, 2000, 2001; Hiltpold et al., 2000), and small proteins of ~60 residues (Gsponer and Caflisch, 2001, 2002). Furthermore, the non-Arrhenius behavior of the temperature-dependence of the folding rate of two structured peptides was demonstrated with the same force field and implicit solvation model (Ferrara et al., 2000). Despite the lack of friction due to the absence of explicit water molecules, the implicit solvent model yields a separation of timescales consistent with experimental data near room temperature: helices fold in ~1 ns (Ferrara et al., 2000; ≈0.1 μs, experimentally, Eaton et al., 2000), β-hairpins in ~10 ns (Ferrara et al., 2000; ≈1 μs, Eaton et al., 2000), and triple-stranded β-sheets in ~100 ns (Cavalli et al., 2002, 2003; ≈10 μs, De Alba et al., 1999).

Denatured conformations

The unfolded state ensemble consists of a large number of different conformers (Cavalli et al., 2003). Hence, it is not possible to sample it at equilibrium either at very high temperature or at physiological temperatures. It was decided to adopt as representative of the denatured state conformations obtained at 550 K to remove any memory of the native state. Indeed, the denaturation of the protein at transition temperature would have shown very long unfolding times. Denatured conformations of the *src*-SH3 domain were obtained by two 6-ns MD simulations at 550 K started from the folded state. Twenty structures were selected with a very low number of native contacts (<5%) and all ω-angles in *trans* conformation.

Characterization of TSE

For a given conformation Γ and amino acid i, a ϕ-value can be approximated (Gsponer and Caflisch, 2002; Li and Daggett, 1994; Vendruscolo et al., 2001) by

$$\phi_i^{calc}(\Gamma) = \frac{N_i(\Gamma)}{N_i^{native}}, \quad (2)$$

where N_i^{native} is the number of contacts of the i^{th} side chain that are present for more than two-thirds of the simulation time of a control 6-ns run at 300 K from the folded state (Gsponer and Caflisch, 2001), and $N_i(\Gamma)$ is the number of native contacts in the conformation Γ. The distance from the experimentally determined TSE for a given conformation Γ is defined using the ϕ_{rmsd} (Gsponer and Caflisch, 2002; Vendruscolo et al., 2001),

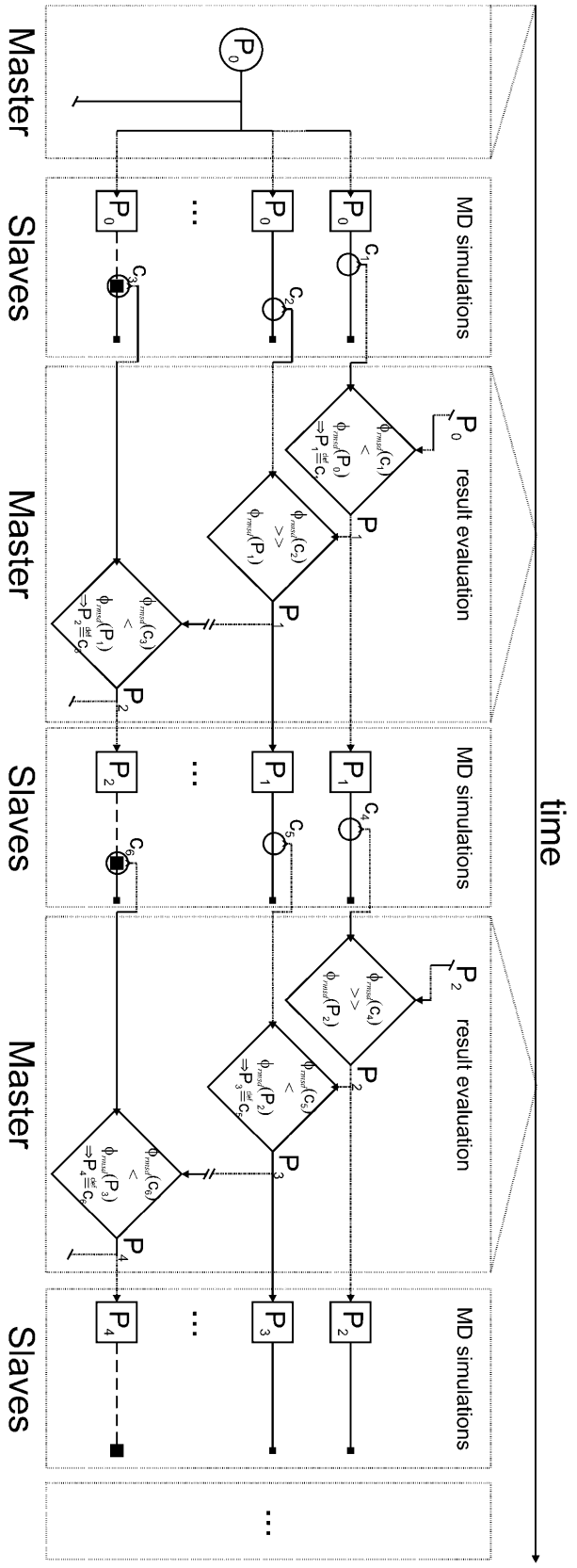

FIGURE 1 Scheme of the LCMD procedure. From left to right, the master node distributes the MD simulations to the slave nodes sending them the initial parent conformation, P_0 (a denatured structure of *src*-SH3). The slave nodes complete the MD simulations started from the parent structure and return the conformation, c_i, with the lowest ϕ_{rmsd} along the trajectory. The master node analyzes one after the other the results from the slave nodes, updates the parent conformation, P_i, when necessary, and sends new MD simulation jobs to the slave nodes. The procedure is repeated until no further improvement of ϕ_{rmsd} of the parent conformations is observed. The optimal trajectory is reported with bold dashed line. The CPU time spent by the master node in the analysis phase is very small with respect to the CPU time spent by the slave nodes to perform the MD simulations.

$$\phi_{\mathrm{rmsd}} = \sqrt{\frac{1}{M_\phi}\sum_i \left(\phi_i^{\mathrm{calc}}(\Gamma) - \phi_i^{\mathrm{exp}}\right)^2}, \qquad (3)$$

where ϕ_i^{exp} is the experimentally determined ϕ-value for residue i, M_ϕ is the number of terms in the sum, and the sum is extended to a selected subset of mainly hydrophobic residues (see below).

Loosely coupled MD simulations (LCMD)

A computational protocol was developed to run multiple MD simulations and continue only those that approach the TSE. An LCMD run (Fig. 1) consists of a sequence of MD simulations starting from an evolving *parent* conformation defined in the following way:

1. The initial *parent* conformation is selected from the pool of denatured conformations.
2. At the end of each MD simulation the conformation with the smallest ϕ_{rmsd} is localized and its ϕ_{rmsd} is compared with the ϕ_{rmsd} of the current parent conformation. The new conformation is accepted and replaces the current parent conformation with a probability that follows a Metropolis-like ansatz,

$$P = \exp\left(-\beta_M\left(\phi_{\mathrm{rmsd}}(\Gamma_{\mathrm{new}}) - \phi_{\mathrm{rmsd}}(\Gamma_{\mathrm{parent}})\right)\right), \qquad (4)$$

where $\beta_M = 0.0005$ is chosen to obtain an arbitrary average acceptance ratio ranging from 0.01 to 0.1 A master/slave protocol is used to distribute the MD simulations to the nodes of a parallel computer. The master node assigns the MD simulations to the slave nodes, evaluates the results from the MD simulations as soon as they are completed by the slave nodes, and updates the parent conformation for the next MD simulations. The LCMD run is continued at least until no further relevant decrease in the ϕ_{rmsd} of the parent conformation is observed over a total simulation time of 150 times the length of the single MD simulation.

The MD simulations are performed at constant temperature T_{eq} using the Berendsen thermostat and a coupling constant of 5.0 ps (Berendsen et al., 1984). Initial velocities are assigned according to a random Maxwellian distribution at T_{ini}. In the present work, T_{eq} was set to 343 K (10 K above the estimated T_m of the *src*-SH3 domain). Two sets of LCMD runs were performed where T_{ini} was set to 343 K and 450 K, respectively. The latter value was used to help escaping from local minima of ϕ_{rmsd}. The time length of the MD simulations in a LCMD run was fixed. Values of 0.02 ns, 0.05 ns, 0.1 ns, 0.2 ns, 1 ns, and 10 ns were used in different LCMD runs. The continuous segments of trajectory leading from the initial parent conformation to the last accepted parent conformation define the LCMD trajectory.

The LCMD runs were performed using two subsets of ϕ_i^{exp} in Eq. 3. The first one (*B*) is the same as in Gsponer and Caflisch (2002), and the second one (*A*) is a modified version of *B* taking into account experimental insights on the importance of particular residues in the folding process, not necessarily directly contained in their ϕ-values. The ϕ-values of the diverging turn were artificially increased by 0.3 units in *A* to take into

account the presence of structure in the denatured state of these residues (Riddle et al., 1999). Moreover, a $\phi^{exp} = 0.8$ for Ile-110 was introduced to take into account the relevance of this residue in the folding nucleus of *fyn*-SH3 domain that has a sequence and a structure very similar to *src*-SH3 (Northey et al., 2002a). Pro-133 was introduced into the ϕ_{rmsd} calculation because it is hydrophobic and buried and has a high ϕ-value (Riddle et al., 1999). Ser-123 and Thr-126 were also added to the ϕ_{rmsd} definition because they form a network of hydrogen bonds in TSE also present in the native state (Grantcharova et al., 2000).

The LCMDs were run on two Beowulf clusters equipped with 64 MP1800+ and 64 MP2100+ Athlon processors, respectively. The number of parallel jobs for each LCMD run ranged from 10 to 30.

Identification of structures close to the TSE

Fifty-four LCMD runs were performed with different starting structures, MD simulation time lengths, and values of T_{ini} for a total of ~20 μs of simulation time. In each simulation the value of the ϕ_{rmsd} parameter was monitored. The final structures of the LCMD trajectories where the ϕ_{rmsd} decreased below a threshold of 0.2 (Gsponer and Caflisch, 2002) and 0.25, for ϕ_{rmsd} subsets *B* and *A*, respectively, constitute an ensemble defined as MDTSE (Table 1), i.e., the ensemble of structures having minimal ϕ_{rmsd} (Gsponer and Caflisch, 2002; Vendruscolo et al., 2001). Most of the residues of these structures have ϕ^{calc} close to ϕ^{exp}.

MD with no bias from MDTSE structures

A total of 176 conventional MD simulations were performed at 310 K and 315 K, starting from selected MDTSE conformations (Table 2). The simulations were 20- or 40-ns long. These simulations allow us to assess the strength of the bias introduced in the LCMD simulations by the loose coupling. Furthermore, they give an estimate of the p_{fold} (Du et al., 1998; Gsponer and Caflisch, 2002) of the starting structures, which is the probability to reach the native conformation before unfolding.

RESULTS

Different extents of ϕ_{rmsd} decrease were observed in the LCMD runs. Eight of them led to MDTSE conformations (Table 1). The other runs led to conformations that represent local minima of the ϕ_{rmsd} from which the system could not escape during the course of the simulations. Some of these conformations show a small fraction of the features of MDTSE but the overall agreement with experimental TSE is low, as evidenced by $\phi^{calc} \neq \phi^{exp}$. Thus, in the following we focus only on the MDTSE conformations. The ϕ_i^{calc} averaged over the eight MDTSE conformations for the 21 mainly hydrophobic residues used in the subset *A* ϕ_{rmsd} definition have a linear correlation coefficient ρ of 0.93 with the corresponding ϕ_i^{exp} (Fig. 2 *a*). The ρ for ϕ_i^{calc} and ϕ_i^{exp} considering 34 residues (both hydrophobic and hydrophilic) is 0.77 (Fig. 2 *b*). The cross-validated correlation coefficient ρ_{cv} for the 13 residues not used in the definition of ϕ_{rmsd} *A* is 0.68. The main outlier is represented by His-122, which is partially solvent-exposed in the x-ray structure of *src*-SH3. Leaving out His-122 results in a $\rho = 0.88$ for 33 residues and $\rho_{cv} = 0.71$ for 12 residues not included in ϕ_{rmsd} calculation. The major deviations from the x-ray structure in the MDTSE ensemble are localized on the C-terminus (residues 134–140) and the N-terminal segment up to the end of the RT-loop (residues 85–105) (Fig. 2 *c*). The central β-sheet (residues 106–133) shows smaller deviations from the x-ray conformation than the other parts of the protein. The RMSD from the x-ray structure of this portion of the protein is 2.3 ± 0.5 Å averaged over the MDTSE ensemble (the values in Fig. 2 *c*

TABLE 1 List of LCMD runs

		Final conformations								Simulation data			
Run	ϕ def.*	ϕ_{rmsd}	$RMSD_0$† (Å)	R_{gyr}‡ (Å)	3-β RMSD§ (Å)	Q¶	Q_{2-3}‖	Q_{3-4}**	S†† (Å^2)	T_{ini}‡‡ (K)	L§§ (ns)	Traj.¶¶ (ns)	Tot.‖‖ (ns)
r-1	B	0.077	10.3	11.1	2.8	0.472	0.917	0.944	181	450	0.1	0.80	200.0
r-2	B	0.156	8.9	11.9	3.4	0.319	0.833	0.667	180	450	0.02	0.12	44.1
r-3	B	0.171	11.9	10.7	6.5	0.236	0.333	0.611	80	450	0.2	12.69	194.0
r-4	A	0.177	8.8	10.1	2.0	0.444	1.000	1.000	117	450	1.0	43.25	405.0
r-5	A	0.181	6.5	10.7	2.0	0.472	1.000	1.000	164	450	1.0	13.01	199.0
r-6	B	0.188	10.2	10.8	4.0	0.222	0.417	0.500	167	450	0.1	0.86	62.8
r-7	B	0.189	11.4	11.0	9.0	0.264	0.333	0.667	86	450	0.05	0.37	215.0
r-8	A	0.244	9.3	11.0	1.9	0.431	1.000	1.000	208	343	0.2	5.50	101.4
r-9***	A	0.331	9.9	11.1	2.0	0.389	0.917	0.833	153	343	0.1	1.50	15.5

*Subset of ϕ^{exp} used in the ϕ_{rmsd}.
†RMSD of the C_α atoms from the x-ray structure.
‡Radius of gyration.
§RMSD of the C_α atoms of residues 106–133 from the x-ray structure.
¶Fraction of C_α native contacts.
‖Fraction of C_α native contacts between strand β2 and β3.
**Fraction of C_α native contacts between strand β3 and β4.
††Buried surface of residues Ile-110, Ale-121, and Ile-132 by contacts between themselves; reference value in native state is 222 Å^2.
‡‡Temperature used to assign the initial random velocities.
§§Length of individual MD simulations.
¶¶Time length of the LCMD trajectory.
‖‖Total simulated time along the LCMD run.
***ϕ_{rmsd} >0.25, but β2-β3-β4 sheet formed.

TABLE 2 MD simulations from MDTSE

I.c.*	$N^{\dagger}$	$L^{\ddagger}$ (ns)	$T^{\S}$ (K)	3-β RMSD¶ (Å)	RMSD$^{\|}$(Å)	n_f**	n_u††
r-1	40	20	310	4.3 ± 0.9	9.3 ± 1.3	16	0
r-1	32	40	315	4.8 ± 1.2	9.3 ± 1.4	19	0
r-2	24	40	315	6.7 ± 1.7	9.6 ± 1.3	5	16
r-4	40	40	310	3.2 ± 0.4	9.3 ± 0.7	0	28
r-5	40	40	310	3.6 ± 0.6	7.6 ± 1.3	0	24

*Initial conformation from MDTSE.
†Number of runs with different initial assignment of random velocities.
‡Run time length.
§Temperature of the run.
¶Average C_α RMSD from x-ray structure ± SD computed along all the runs for residues 106–133.
‖Average C_α RMSD from x-ray structure ± SD computed along all the runs for all residues.
**Number of runs where $\langle RMSD\rangle < RMSD_0 - \sigma_{RMSD}$; σ_{RMSD} is the SD of the RMSD along the run and $RMSD_0$ is the C_α RMSD of the initial conformation.
††Number of runs where $\langle RMSD\rangle > RMSD_0 + \sigma_{RMSD}$; notation as above.

for the $\beta2$-$\beta3$-$\beta4$ are larger because the superposition is optimized over all of the protein).

The analysis of the native contacts of MDTSE conformations reveals that the native topology of the three-stranded β-sheet is reproduced (Table 1). In particular, in four of them, r-1, r-4, r-5, and r-8, the secondary structure of this β-sheet is fully formed (Fig. 3), according to STRIDE (Frishman and Argos, 1995). In agreement with ϕ-value analysis and previous observations for the TSE of *src*-SH3 (Gsponer and Caflisch, 2002) the residues of the *n-src*-loop form tighter interactions in MDTSE conformations than in the native state as seen in r-1, r-4, r-5, and r-8, where strand $\beta2$ is elongated. Moreover, the packing of the hydrophobic residues Ile-110, Ala-121, and Ile-132 is almost native-like (Table 1) and the N-terminal segment of the protein up to the end of RT-loop and the C-terminus are unstructured. Three of the MDTSE structures obtained using subset *B* in ϕ_{rmsd} calculation present relatively large RMSD for the central β-strand with respect to the x-ray structure and two of them show also a relatively looser packing of the nucleus residues than the native state (Table 1). In these structures the $\beta2$-$\beta3$ strand has more than one-half of the native contacts formed, whereas the $\beta3$-$\beta4$ strand is less structured. These features are in contrast with data from glycine loop insertion and disulphide cross-link experiments (Grantcharova et al., 2000) and from experiments revealing the detailed packing

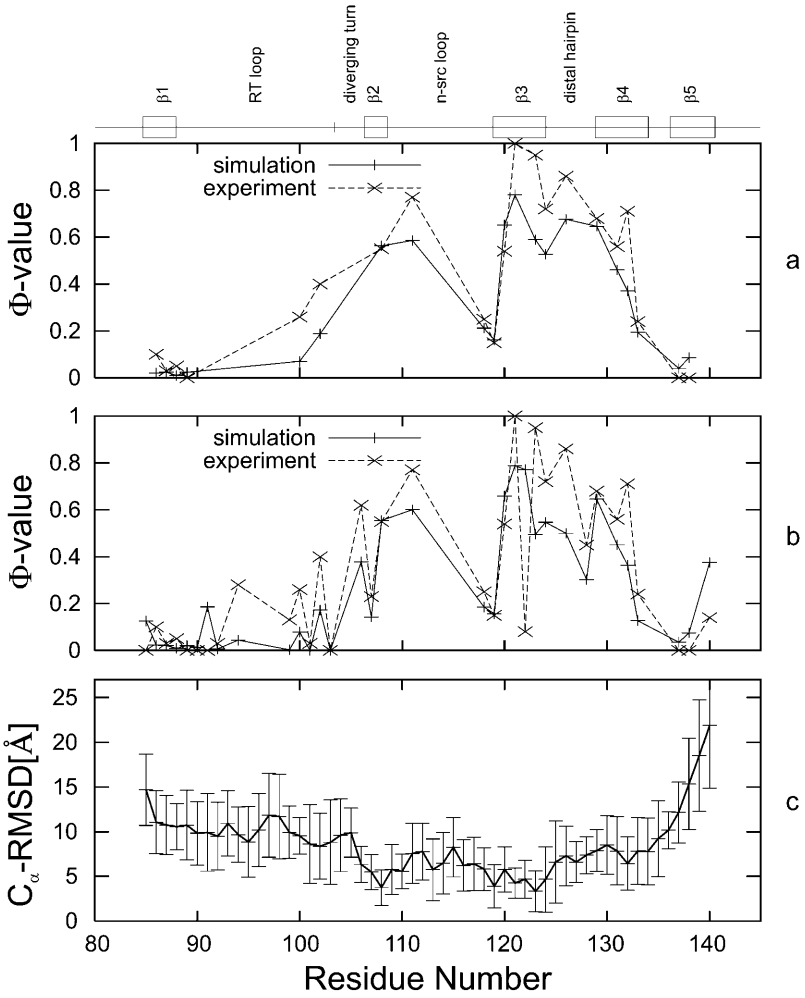

FIGURE 2 Average properties of the MDTSE. (*a*) ϕ^{calc} (*continuous line*) and ϕ^{exp} (*dashed line*) profile for the 21 residues used in ϕ_{rmsd} *A* calculation. (*b*) ϕ^{calc} (*continuous line*) and ϕ^{exp} (*dashed line*) profile for 34 residues with known ϕ^{exp}. (*c*) C_α mean deviation from x-ray structure, after structural superposition. The error bar represents the standard deviation computed along the MDTSE set.

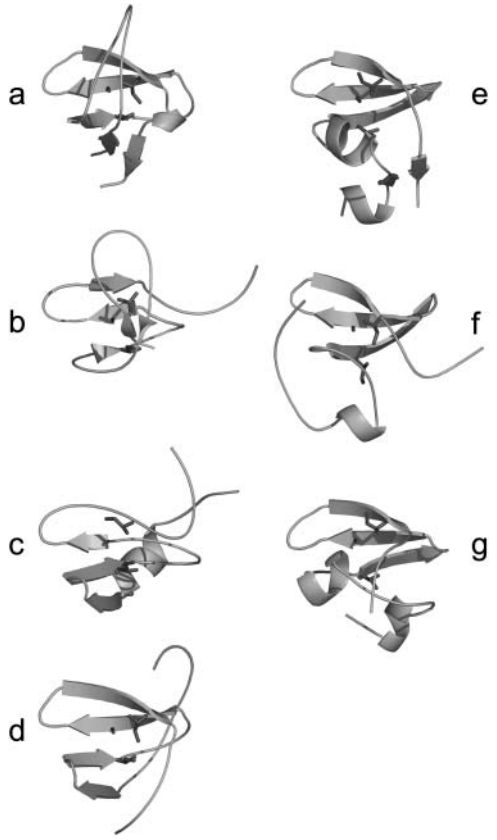

FIGURE 3 (*a*) Native structure of the *src*-SH3 domain. Final conformations of *b*, r-2; *c*, r-9; *d*, r-8; *e*, r-5; *f*, r-1; and *g*, r-4. Secondary structure is represented by cartoons and residues of the folding nucleus Ile-110, Ala-121, and Ile-132 are in sticks. This figure was made with PyMOL (DeLano, 2002).

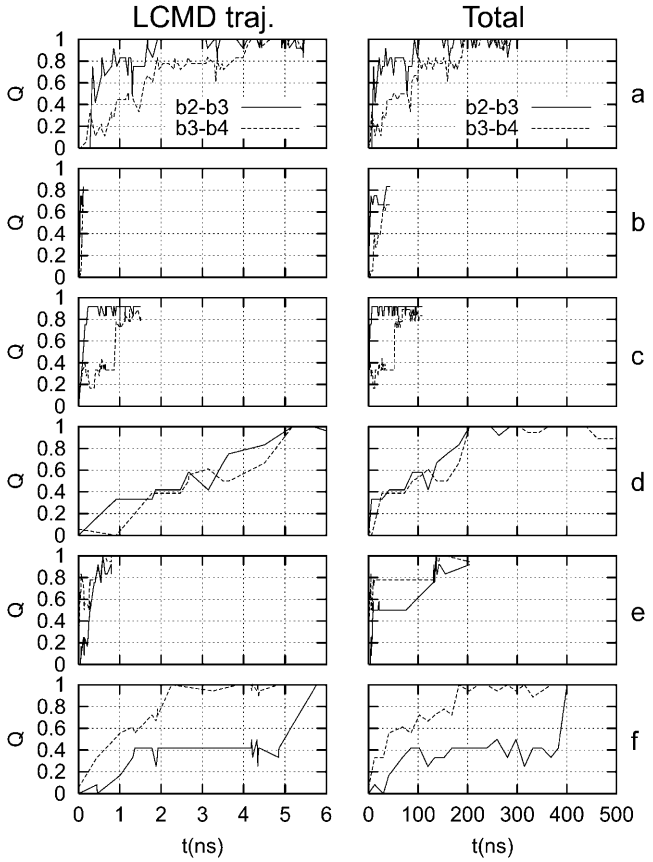

FIGURE 4 Folding pathways of the $\beta2$-$\beta3$-$\beta4$ sheet of *src*-SH3 along the LCMD runs leading to MDTSE conformations. The continuous line represents the fraction of native contact in $\beta2$-$\beta3$ hairpin and the dashed line represents the fraction of contacts in the $\beta3$-$\beta4$ hairpin. The time on the *x* axis in the left column represents the LCMD trajectory time and the right column represents total simulated time. *a*, r-8; *b*, r-2; *c*, r-9; *d*, r-5; *e*, r-1; and *f*, r-4. MD simulations time length is 0.2, 0.02, 0.1, 1.0, 0.1, and 1.0 ns, respectively.

of core hydrophobic residues (Northey et al., 2002a). Thus, although the subset *B* was sufficient to pinpoint TSE structures along the unfolding simulations (Gsponer and Caflisch, 2002), it represents a loose criterion for the selection of TSE conformations in a larger and less homogeneous set of structures where it showed itself to be less efficient. The subset *A* resulted to be more appropriate for the present study.

The LCMD trajectories leading to the MDTSE conformations describe the pathways followed by *src*-SH3 approaching the TSE from the denatured state. The ϕ_{rmsd} decreases along these pathways according to the definition of the LCMD procedure. Along each LCMD trajectory the RMSD of the residues in the central three-stranded β-sheet (residues 106–133) has a high correlation with the ϕ_{rmsd} (correlation coefficient of 0.8) in agreement with the fact that the TSE of *src*-SH3 domain is mainly structured in the central three-stranded β-sheet (Riddle et al., 1999). There are three possible pathways to the folding of the three-stranded β-sheet: 1), formation of $\beta2$-$\beta3$ hairpin followed by formation of $\beta3$-$\beta4$ hairpin; 2), the inverted sequence of events; and 3), concomitant formation of the two hairpins (Fig. 4). The formation of the $\beta3$-$\beta4$ hairpin proceeds from the proximal contacts to the distal ones (Fig. 5). On the other hand, the $\beta2$-$\beta3$ hairpin forms more cooperatively by native interactions with a contact order between 9 and 15 and non-native contacts with lower contact order (Fig. 5). Then, native interactions with low contact order gradually form (Fig. 5), but a fraction of non-native interactions with low contact order remains. The latter corresponds to the contacts between the C-terminal part of the elongated strand $\beta2$ and the N-terminal part of $\beta3$.

The LCMD trajectories leading to the complete formation of the three-stranded β-sheet have significantly different time lengths, whereas the total amount of simulated time needed to observe the formation of the β-sheet, that is the LCMD trajectory length plus the sum of the lengths of the discarded segments of trajectories, varies between 100 and 400 ns. These small variations indicate that the rate of formation of the β-sheet does not depend on the MD simulation lengths used in the LCMD run. However, the absence of a frictional

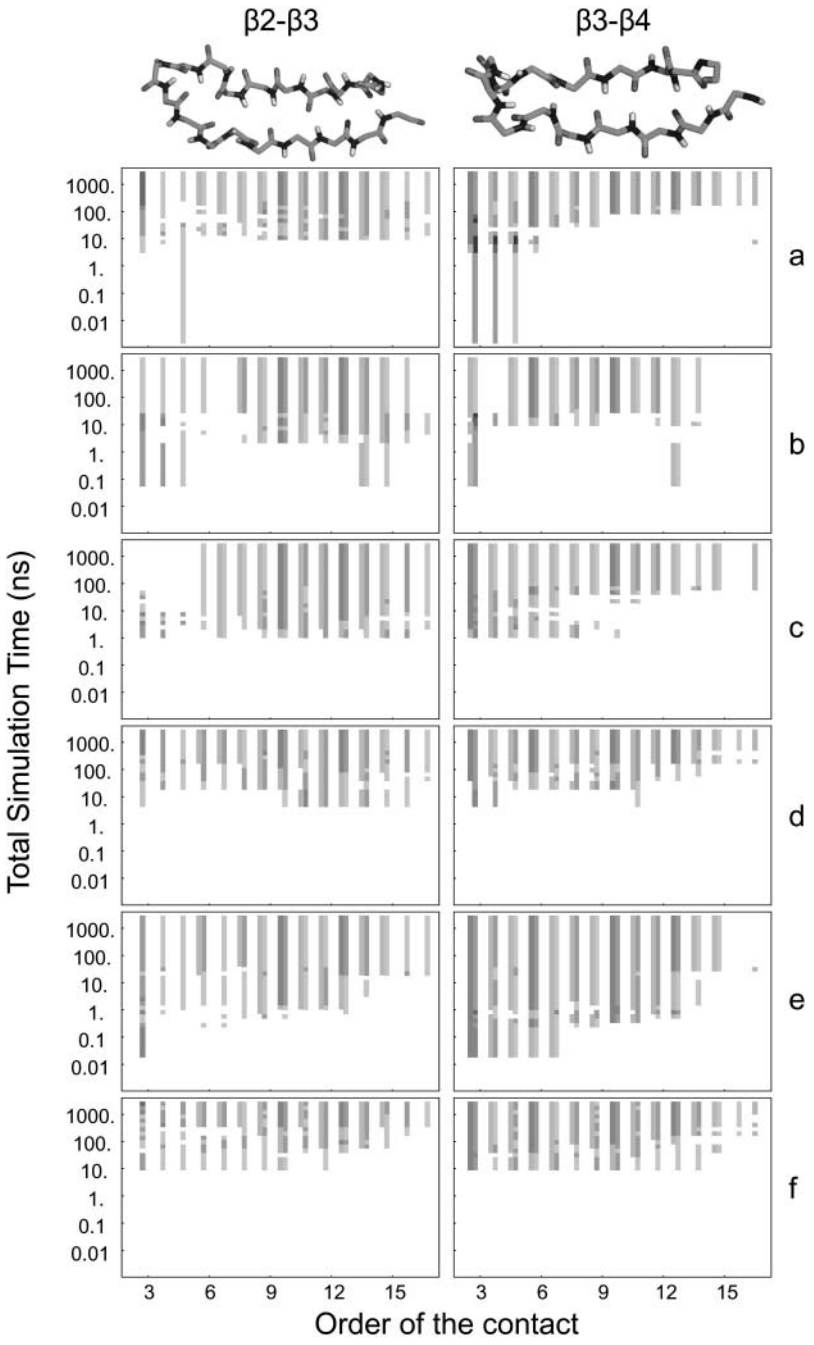

FIGURE 5 The detailed formation of the $\beta 2$-$\beta 3$ (*left column*) and $\beta 3$-$\beta 4$ (*right column*) hairpins along the LCMD runs leading to MDTSE conformations. The colored bars represent the number of contacts with the specified contact order formed at the specified time (the *red* intensity scale ranges from one to five native contacts and the *gray* scale from one to five non-native contacts). *a*, r-8; *b*, r-2; *c*, r-9; *d*, r-5; *e*, r-1; and *f*, r-4. MD simulations time length is 0.2, 0.02, 0.1, 1.0, 0.1, and 1.0 ns, respectively.

term in the simulations does not allow us to make a reliable estimate of folding rates.

The main purpose of the present work was to investigate the early events of folding. Nevertheless, to test the robustness of the approach two LCMD runs were started from the native structure. In the final conformations obtained from these simulations, the average RMSD from x-ray structure, radius of gyration, and fraction of native contacts (3.6 Å, 10.7 Å, and 0.6, respectively) show that they are similar to the TSE conformations obtained from unfolding simulations presented in Gsponer and Caflisch (2002). Since these LCMD simulations were started from the native state, the final conformations are more native-like than the MDTSE. The agreement with ϕ^{exp} is as good as for MDTSE (average set A $\phi_{\mathrm{rmsd}} = 0.07$). These simulation results suggest that the ensemble that minimizes the ϕ_{rmsd} (and, thus, is in agreement with TSE determined by ϕ^{exp}) spans conformations with different degrees of nativeness.

In 40 of the conventional MD simulations started from MDTSE conformations (Table 2) the average C_α-RMSD from the x-ray structure decreased >1 SD below the initial value. On the contrary, in 68 simulations this value increased

1 SD above the initial value. In the 68 remaining simulations, the average RMSD did not significantly increase or decrease. In particular, in the simulations from r-4 and r-5, the $\beta2$-$\beta3$-$\beta4$ sheet was conserved on average whereas the other parts of the protein acquired a more native-like conformation. More fluctuations were seen in the other simulations. In 95% of the simulations from r-5 and in 25% of the simulations from r-1 the total RMSD from the x-ray structure reached values <6.0 Å and in some cases between 5.0 and 4.0 Å. Notwithstanding the fact that a fraction of the trajectories approached the native conformation, none of them reached the folding conditions as defined in Gsponer and Caflisch (2002; i.e., $RMSD < 2.5$ Å and $Q > 0.875$). As a comparison, an average RMSD of 2.9 ± 0.2 Å was observed in a 40-ns 300-K simulation started from the x-ray structure.

DISCUSSION

Eight LCMD trajectories of the *src*-SH3 domain started from the unfolded state reached conformations showing several characteristics of the TSE. These characteristics include: $\phi^{calc} \approx \phi^{exp}$ (Riddle et al., 1999), formation of the $\beta2$-$\beta3$-$\beta4$ sheet (Grantcharova et al., 2000; Riddle et al., 1999), packing of the folding nucleus (Northey et al., 2002a), and disordered N- and C-termini and RT-loop. The ϕ-values are reproduced with high accuracy even for most of the residues that were not used in the definition of ϕ_{rmsd}. The degree of accuracy is similar to that previously achieved along unfolding trajectories at 375 K (Gsponer and Caflisch, 2002). The $\beta2$-$\beta3$-$\beta4$ sheet is formed and its structure is very close to the x-ray conformation. Furthermore, in most of the MDTSE conformations $\beta2$ is elongated in a tight hairpin in agreement with previous simulation results (Gsponer and Caflisch, 2002). The $\beta2$ elongation is consistent with experimental ϕ-values >1.0 in the *n-src* loop region that may be due to non-native contacts made by those residues in TSE (see below). The packing of the residues forming the folding nucleus of SH3 domain (Northey et al., 2002a) resembles the native-like packing . The RT-loop and N- and C-termini of these structures are mainly unstructured in agreement with their low ϕ-values. The disorder in these segments may be overestimated with respect to TSE conformations, as suggested by the comparison of the experimental entropy loss after the disulphide cross-link of the RT-loop base with theoretical estimations that reveal non-random-coil structure in denatured RT-loop (Grantcharova et al., 2000); however, those data are not conclusive regarding this point. A comparison of MDTSE with the TSE structures obtained from unfolding simulations (Gsponer and Caflisch, 2002) reveals that the former present larger deviations from the x-ray conformation than the latter. This is not surprising since the LCMD runs were started from denatured state whereas the simulations in Gsponer and Caflisch (2002) were started from the x-ray structure. The main discrepancies are concentrated in the contacts between the two branches of the RT-loop, the tip of the RT-loop, and the central β-strand, and the contacts between N- and C-termini, whereas the $\beta2$-$\beta3$-$\beta4$ sheet is preserved in both sets at a similar extent (Fig. 6). In most of the TSE structures

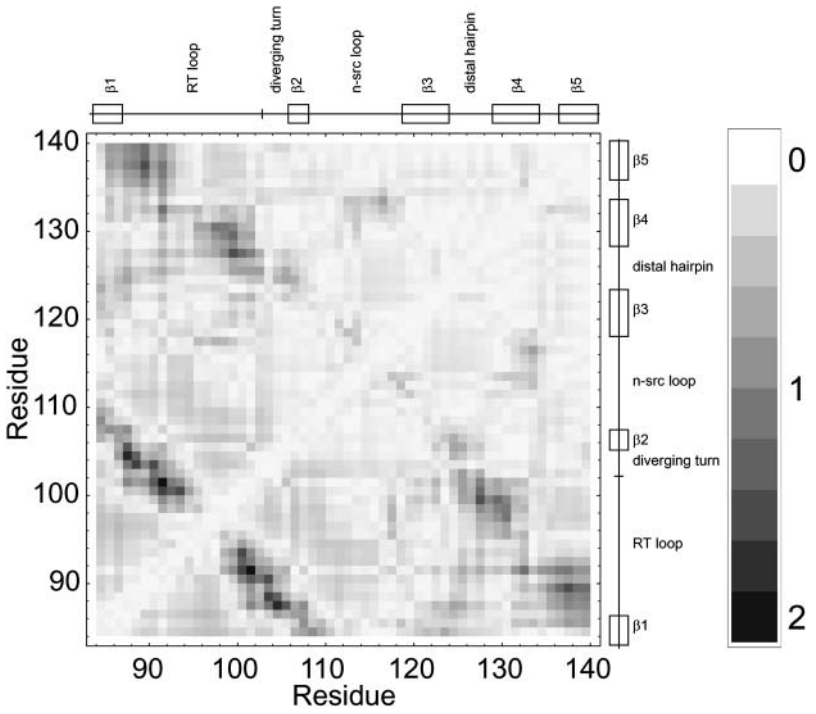

FIGURE 6 The absolute value of the relative difference between the average distance of pairs of residues in the MDTSE structures and in the putative TSE structures from Gsponer and Caflisch (2002). Large relative differences (*dark*) are concentrated in the RT-loop, N-, and C-termini, and contacts between the tip of the RT-loop and the central β-sheet.

from Gsponer and Caflisch (2002), the native topology is also completely preserved in the N- and C-termini, the RT-loop, and the contacts between the central β-strand and the tip of the RT-loop, whereas in the MDTSE structures the native topology is present only in the $\beta2$-$\beta3$-$\beta4$ sheet. These deviations between the two ensembles occur even if the overall deviation from the experimental data, defined by ϕ_{rmsd}, is very similar. Recent biased sampling simulations of *src*-SH3 with explicit water (Guo et al., 2003) identify a poorly native-like TSE for the rate-limiting step of the folding transition; in that case the number of native contacts is on average 0.3 and the radius of gyration is larger than that of MDTSE conformations. Although the small discrepancy in the number of native contacts may be determined by the slightly different definition of native contact, the different degree of compactness of the structures is probably due to the two solvation methods. These differences, however, involve mainly the flexible parts of the TSE (i.e., the RT-loop and the termini), whereas good agreement holds on the structured parts (i.e., central three-stranded β-sheet).

The LCMD procedure allows us to obtain insights into the sequence of events along the folding pathway of *src*-SH3. The $\beta2$-$\beta3$ hairpin and the $\beta3$-$\beta4$ hairpin form before the other parts of the protein. This finding is in agreement with experimental data from ϕ-value analysis (Grantcharova et al., 1998), disulphide cross-linking, and glycine-loop insertion experiments (Grantcharova et al., 2000) and validates the results from high temperature unfolding simulations with implicit (Gsponer and Caflisch, 2001) and explicit (Tsai et al., 1999) treatment of the solvent. The latter is not always true, since folding and unfolding pathways may not coincide when a large perturbation is applied to the system (like the high temperature used to obtain a fast unfolding; Dinner and Karplus, 1999b; Shea and Brooks, 2001). Folding and unfolding pathways of *src*-SH3 are essentially the same (time-inverted) as already pointed out in Shea et al. (2002) probably because of the particularly polarized structure in the TSE. This result is also in agreement with the thermodynamic picture obtained by importance-sampling molecular dynamics (Shea et al., 2002). The diversity of the pathways observed for the formation of the central β-sheet matches the diversity observed by high temperature MD unfolding (Gsponer and Caflisch, 2001); however, the number of pathways that has been collected in the present work is not enough to reliably assess the relative weight of each pathway. The presence of a similar branching in the folding pathways of proteins has been experimentally measured in the case of protein G and protein L at the rate-limiting step of the folding reaction (McCallister et al., 2000) and more recently for the denaturant-induced unfolding of titin (Wright et al., 2003).

The two hairpins form in different ways: in the case of the $\beta3$-$\beta4$ hairpin, the formation of local contacts at its tip, which are favorable for entropic reasons, precedes those at the distal end of the hairpin. This picture is not in contrast with the proposed looping of the $\beta3$-$\beta4$ hairpin in its middle (Grantcharova et al., 2000; i.e., disorder in the packing of side chains in the middle of the hairpin), because the progressive ordering of the backbone is followed by a partial ordering of the corresponding side chains that is compatible with the experimental ϕ-values (Grantcharova et al., 1998). On the other hand, the $\beta2$-$\beta3$ hairpin collapses more cooperatively than $\beta3$-$\beta4$ even if the earliest proximal contacts are non-native. The native contacts close to the tip (i.e., the native *n-src* loop) progressively substitute the non-native contacts after the initial collapse. In this case the energy decrease due to the early hydrophobic collapse of Leu-108, Ile-110, and Val-111 on strand $\beta2$, and Leu-120 and Ala-121 on strand $\beta3$, as opposed to the entropy increase, may lead the process. The looping of this hairpin is evident even at the backbone contact level; however, it takes place at the tip and not in the middle of the hairpin, like in the $\beta3$-$\beta4$ (see above and Grantcharova et al., 2000). The looping in the middle of the hairpin was also observed for one of the hairpins in protein L (Kim et al., 2000) and has been proposed as a common theme in protein folding (Grantcharova et al., 2000). The present results show that the looping may also occur at the tip of the hairpin and non-native contacts may play a relevant role in that case. The importance of the non-native contacts in initiating the formation of the $\beta2$-$\beta3$ hairpin could explain the experimental ϕ-values >1.0 (Grantcharova et al., 1998) for many of the residues of the *n-src* loop and agrees with the suggestion that non-native contacts may speed up the folding process (Li et al., 2000).

A detailed description of the formation of the folding nucleus of the *src*-SH3, that is, the precritical phase of the folding process, has been addressed in this work. Clearly, validation is needed and this might require novel experimental approaches (Schymkowitz et al., 2002).

The role of the constraints in determining the pathways to TSE can be inferred by comparing the results obtained with the two sets *A* and *B*. Set *A* contains constraints on the ϕ-value of 22 residues whereas set *B* only involves 18 residues. Also, the ϕ_{exp} of residues in the diverging turn of set *A* have been increased implying a further increase of the number of constraints (i.e., more native contacts are required to lower the ϕ_{rmsd}). Some LCMD runs with set *B* reached false-positive TSE structures (i.e., structures where important features of the TSE are not fully present, as shown in the Results section); yet, the runs that reached TSE structures consistent with experimental data (r-1 and r-2, see Table 1) show patterns of formation of the folding nucleus of *src*-SH3 similar to those obtained with set *A* (Figs. 4 and 5). In addition, the length of the single MD simulations in LCMD runs, that is, the frequency used to enforce the constraints along the run, does not affect these patterns or the total simulation time needed for the formation of the three-stranded β-sheet (Fig. 4). From this comparison, it would emerge that the constraints only affect the efficiency of the

LCMD to reach TSE conformations, whereas the main characteristics of the observed pathways seem to be independent from them.

The presence of parallel microscopic flow processes has also been indicated as a possible explanation of abnormal ϕ-values (Ozkan et al., 2001). At the light of such interpretation of our data, the mutation of residues in the *n-src* loop with abnormal ϕ-values may shift the equilibrium wild-type sequence of formation of the two β-hairpins in one direction or the other, favoring, for example, the early formation of $\beta2$-$\beta3$ and a later formation of the $\beta3$-$\beta4$ hairpin. Simulations with mutants of the *src*-SH3 domain would allow us to rule out this hypothesis, but this is beyond the scope of the present work. The different packing of residues in the native and in the transition state is another possible explanation of the abnormal ϕ-values (Northey et al., 2002b).

The conventional MD simulations started from the MDTSE conformations show that the central β-sheet, characteristic of the TSE ensemble, does not vanish as the pressure on the decrease of the ϕ_{rmsd} is released. This allows us to conclude that the bias introduced through the LCMD procedure does not strongly affect the dynamics of the protein by forcing the sampling of very high free energy regions. These and the p_{fold} data may also indicate that the saddlepoint in the free energy landscape representing the TSE of this protein with this force field is quite flat, and does not allow the system to fold or unfold within the timescales that have been sampled in the present work—similar to findings in the case of acylphosphatase (Paci et al., 2002). This fact is not in disagreement with the results presented in Gsponer and Caflisch (2002) because in that case very strict conditions were assumed to define the unfolded state (i.e., $RMSD > 7.0$ Å and $Q < 0.375$). Indeed, these conditions are already met by several of the MDTSE structures, whereas during the unbiased MD simulations that started from these conformations the trajectories approach the native structure passing through those thresholds. The simulation results indicate that, within the region of phase space where ϕ_{rmsd} is low, LCMD allows us to reach locally more stable conformations (i.e., with higher diffusion times) than unfolding simulations. The different p_{fold} of MDTSE and TSE structures obtained from unfolding simulations could also depend on the force field that might not be accurate enough to reproduce the p_{fold} of the analyzed conformations (Paci et al., 2002). Statistical variations may also determine this difference; however, 10 simulations that clearly reach well-defined unfolded or folded states like in Gsponer and Caflisch (2002) are sufficient to estimate p_{fold} with a maximal standard deviation of the order of 0.16 as computed from a binomial distribution.

Compared to the distributed computing approach to protein folding pioneered by Pande's group (Shirts and Pande, 2001; Snow et al., 2002), the LCMD procedure presents a key difference that addresses the major problem inherent to that method (Fersht, 2002; Paci et al., 2003). In LCMD, only simulations that approach the experimentally determined transition state are continued, so that the system has a minor chance of exploring minor pathways and statistically irrelevant regions of the conformational space. On the other hand, LCMD is possible only if the TSE of the system under study has been experimentally probed with ϕ-value analysis. The diversity found in the folding pathways of the central β-sheet of the *src*-SH3 domain offers an a posteriori validation of the LCMD procedure; different folding pathways are explored but they reach TSE conformations consistent with experimental data. Furthermore, the independence of the rate of central β-sheet formation from the time length of the individual MD simulations indicates that the loose coupling of simulations does not apparently affect the LCMD runs.

G.S. thanks M. Vendruscolo for interesting discussions and E. Paci for critical reading of the manuscript. The simulations were performed on a Beowulf cluster running Linux, and we thank U. Haberthür, A. Cavalli, and F. Rao for their invaluable help in setting up the cluster and computer support. We thank A. Widmer (Novartis Pharma, Basel) for providing the molecular modeling program Wit!P, which was used for visual analysis of the trajectories.

This work was supported by the Swiss National Competence Center in Structural Biology and the Swiss National Science Foundation (grant No. 31-64968.01 to A.C.). J.G. is a fellow of the Swiss MD-PhD program (grant No. 3236-057617).

REFERENCES

Berendsen, H. J. C., J. P. M. Postma, W. F. Van Gunsteren, A. Dinola, and J. R. Haak. 1984. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81:3684–3690.

Borreguero, J. M., N. V. Dokholyan, S. V. Buldyrev, E. I. Shakhnovich, and H. E. Stanley. 2002. Thermodynamics and folding kinetics analysis of the SH3 domain from discrete molecular dynamics. *J. Mol. Biol.* 318:863–876.

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

Cavalli, A., P. Ferrara, and A. Caflisch. 2002. Weak temperature dependence of the free energy surface and folding pathways of structured peptides. *Proteins Struct. Funct. Gen.* 47:305–314.

Cavalli, A., U. Haberthur, E. Paci, and A. Caflisch. 2003. Fast protein folding on downhill energy landscape. *Protein Sci.* 12:1801–1803.

Cheung, M. S., A. E. Garcia, and J. N. Onuchic. 2002. Protein folding mediated by solvation: water expulsion and formation of the hydrophobic core occur after the structural collapse. *Proc. Natl. Acad. Sci. USA.* 99:685–690.

Daggett, V., and A. R. Fersht. 2003. Is there a unifying mechanism for protein folding? *Trends Biochem. Sci.* 28:18–25.

Daggett, V., A. Li, L. Itzhaki, D. Otzen, and A. Fersht. 1996. Structure of the transition state for folding of a protein derived from experiment and simulation. *J. Mol. Biol.* 257:430–440.

De Alba, E., J. Santoro, M. Rico, and M. A. Jiménez. 1999. De novo design of a monomeric three-stranded antiparallel β-sheet. *Protein Sci.* 8:854–865.

DeLano, W. L. 2002. The PyMOL Molecular Graphics System. DeLano Scientific, San Carlos, CA.

Ding, F., N. V. Dokholyan, S. V. Buldyrev, H. E. Stanley, and E. I. Shakhnovich. 2002. Direct molecular dynamics observation of protein folding transition state ensemble. *Biophys. J.* 83:3525–3532.

Dinner, A., and M. Karplus. 1999a. The thermodynamics and kinetics of protein folding: a lattice model analysis of multiple pathways with intermediates. *J. Phys. Chem. B.* 103:7976–7994.

Dinner, A. R., and M. Karplus. 1999b. Is protein unfolding the reverse of protein folding? A lattice simulation analysis. *J. Mol. Biol.* 292:403–419.

Du, R., V. S. Pande, A. Y. Grosberg, T. Tanaka, and E. S. Shakhnovich. 1998. On the transition coordinate for protein folding. *J. Chem. Phys.* 108:334–350.

Eaton, W. A., V. Munoz, S. J. Hagen, G. S. Jas, L. J. Lapidus, E. R. Henry, and J. Hofrichter. 2000. Fast kinetics and mechanisms in protein folding. *Ann. Rev. Biophys. Biophys. Struct.* 29:327–359.

Elcock, A. H., and J. A. McCammon. 1997. Continuum solvation model for studying protein hydration thermodynamics at high temperatures. *J. Phys. Chem. B.* 101:9624–9634.

Ferrara, P., J. Apostolakis, and A. Caflisch. 2000. Thermodynamics and kinetics of folding of two model peptides investigated by molecular dynamics simulations. *J. Phys. Chem. B.* 104:5000–5010.

Ferrara, P., J. Apostolakis, and A. Caflisch. 2002. Evaluation of a fast implicit solvent model for molecular dynamics simulations. *Proteins.* 46:24–33.

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

Ferrara, P., and A. Caflisch. 2001. Native topology or specific interactions: what is more important for peptide folding? *J. Mol. Biol.* 306:837–850.

Fersht, A. R. 2002. On the simulation of protein folding by short time scale molecular dynamics and distributed computing. *Proc. Natl. Acad. Sci. USA.* 99:14122–14125.

Frishman, D., and P. Argos. 1995. Knowledge-based protein secondary structure assignment. *Proteins.* 23:566–579.

Grantcharova, V. P., D. S. Riddle, and D. Baker. 2000. Long-range order in the *src*-SH3 folding transition state. *Proc. Natl. Acad. Sci. USA.* 97:7084–7089.

Grantcharova, V. P., D. S. Riddle, J. V. Santiago, and D. Baker. 1998. Important role of hydrogen bonds in the structurally polarized transition state for folding of the *src*-SH3 domain. *Nat. Struc. Biol.* 5:714–720.

Gsponer, J., and A. Caflisch. 2001. Role of native topology investigated by multiple unfolding simulations of four SH3 domains. *J. Mol. Biol.* 309:285–298.

Gsponer, J., and A. Caflisch. 2002. Molecular dynamics simulations of protein folding from the transition state. *Proc. Natl. Acad. Sci. USA.* 99:6719–6724.

Guo, W. H., S. Lampoudi, and J. E. Shea. 2003. Posttransition state desolvation of the hydrophobic core of the *src*-SH3 protein domain. *Biophys. J.* 85:61–69.

Hasel, W., T. F. Hendrickson, and W. C. Still. 1988. A rapid approximation to the solvent accessible surface areas of atoms. *Tetrahedron Comput. Methodol.* 1:103–116.

Hiltpold, A., P. Ferrara, J. Gsponer, and A. Caflisch. 2000. Free energy surface of the helical peptide $Y(MEARA)_6$. *J. Phys. Chem. B.* 104: 10080–10086.

Jackson, S. E., and A. R. Fersht. 1991. Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition. *Biochemistry.* 30:10428–10435.

Kim, D. E., C. Fisher, and D. Baker. 2000. A breakdown of symmetry in the folding transition state of protein l. *J. Mol. Biol.* 298:971–984.

Lazaridis, T., and M. Karplus. 1999. Effective energy function for proteins in solution. *Proteins Struct. Funct. Gen.* 35:133–152.

Li, A. J., and V. Daggett. 1996. Identification and characterization of the unfolding transition state of chymotrypsin inhibitor 2 by molecular dynamics simulations. *J. Mol. Biol.* 257:412–429.

Li, A. J., and V. Daggett. 1994. Characterization of the transition state of protein unfolding by use of molecular dynamics—chymotrypsin inhibitor-2. *Proc. Natl. Acad. Sci. USA.* 91:10430–10434.

Li, L., L. A. Mirny, and E. I. Shakhnovich. 2000. Kinetics, thermodynamics and evolution of non-native interactions in a protein folding nucleus. *Nat. Struct. Biol.* 7:336–342.

Matouschek, A., J. T. Kellis, Jr., L. Serrano, and A. R. Fersht. 1989. Mapping the transition state and pathway of protein folding by protein engineering. *Nature.* 340:122–126.

Mayor, U., N. R. Guydosh, C. M. Johnson, J. G. Grossmann, S. Sato, G. S. Jas, S. M. V. Freund, D. O. V. Alonso, V. Daggett, and A. R. Fersht. 2003. The complete folding pathway of a protein from nanoseconds to microseconds. *Nature.* 421:863–867.

McCallister, E. L., E. Alm, and D. Baker. 2000. Critical role of β-hairpin formation in protein-G folding. *Nat. Struct. Biol.* 7:669–673.

Northey, J. G. B., A. Di Nardo, and A. R. Davidson. 2002a. Hydrophobic core packing in the SH3 domain folding transition state. *Nat. Struct. Biol.* 9:126–130.

Northey, J. G. B., K. L. Maxwell, and A. R. Davidson. 2002b. Protein folding kinetics beyond the ϕ-value: using multiple amino acid substitutions to investigate the structure of the SH3 domain folding transition state. *J. Mol. Biol.* 320:389–402.

Ozkan, S. B., I. Bahar, and K. A. Dill. 2001. Transition states and the meaning of ϕ-values in protein folding kinetics. *Nat. Struct. Biol.* 8:765–769.

Paci, E., A. Cavalli, M. Vendruscolo, and A. Caflisch. 2003. Analysis of the distributed computing approach applied to the folding of a small β-peptide. *Proc. Natl. Acad. Sci. USA.* 100:8217–8222.

Paci, E., M. Vendruscolo, C. M. Dobson, and M. Karplus. 2002. Determination of a transition state at atomic resolution from protein engineering data. *J. Mol. Biol.* 324:151–163.

Riddle, D. S., V. P. Grantcharova, J. V. Santiago, E. Alm, I. Ruczinski, and D. Baker. 1999. Experiment and theory highlight role of native state topology in SH3 folding. *Nat. Struct. Biol.* 6:1016–1024.

Schymkowitz, J. W. H., F. Rousseau, and L. Serrano. 2002. Surfing on protein folding energy landscapes. *Proc. Natl. Acad. Sci. USA.* 99: 15846–15848.

Shea, J. E., and C. L. Brooks. 2001. From folding theories to folding proteins: a review and assessment of simulation studies of protein folding and unfolding. *Annu. Rev. Phys. Chem.* 52:499–535.

Shea, J. E., J. N. Onuchic, and C. L. Brooks. 2002. Probing the folding free energy landscape of the *src*-SH3 protein domain. *Proc. Natl. Acad. Sci. USA.* 99:16064–16068.

Shirts, M. R., and V. S. Pande. 2001. Mathematical analysis of coupled parallel simulations. *Phys. Rev. Lett.* 86:4983–4987.

Snow, C. D., N. Nguyen, V. S. Pande, and M. Gruebele. 2002. Absolute comparison of simulated and experimental protein-folding dynamics. *Nature.* 420:102–106.

Tsai, J., M. Levitt, and D. Baker. 1999. Hierarchy of structure loss in MD simulations of *src*-SH3 domain unfolding. *J. Mol. Biol.* 291:215–225.

Vendruscolo, M., E. Paci, C. M. Dobson, and M. Karplus. 2001. Three key residues form a critical contact network in a protein folding transition state. *Nature.* 409:641–645.

Wright, C. F., K. Lindorff-Larsen, L. G. Randles, and J. Clarke. 2003. Parallel protein-unfolding pathways revealed and mapped. *Nat. Struct. Biol.* 10:658–662.

Xu, W. Q., S. C. Harrison, and M. J. Eck. 1997. Three-dimensional structure of the tyrosine kinase *c-src*. *Nature.* 385:595–602.