PROTEINS: Structure, Function, and Bioinformatics 64:178–192 (2006)

Unfolding Transition State and Intermediates of the Tumor Suppressor $p16^{INK4a}$ Investigated by Molecular Dynamics Simulations

Gianluca Interlandi, Giovanni Settanni, and Amedeo Caflisch*
Department of Biochemistry, University of Zürich, Zürich, Switzerland

ABSTRACT **The ankyrin repeat is one of the most common protein motifs and is involved in protein–protein interactions. It consists of 33 residues that assume a β-hairpin helix-loop-helix fold. Mutagenesis and kinetic experiments (Φ-value analysis of the folding transition state) have shown that the tumor suppressor $p16^{INK4a}$, a four-repeat protein, unfolds sequentially starting from the two N-terminal repeats. Here, the flexibility of $p16^{INK4a}$ at room temperature and its unfolding mechanism at high temperature have been investigated by multiple molecular dynamics runs in explicit water for a total simulation time of 0.65 μs. The transition state ensemble (TSE) of $p16^{INK4a}$ was identified by monitoring both the deviation from the experimental Φ values and sudden conformational changes along the unfolding trajectories. Conformations in the TSE have a mainly unstructured second repeat whereas the other repeats are almost completely folded. A rigid-body displacement of the first repeat involving both a rotation and translation is observed in all molecular dynamics simulations at high temperature. The Trp^{15}, Pro^{75}, and Ala^{76} sidechains are more buried in the TSE than the native state. The sequential unfolding starting at the second repeat is in agreement with the mutagenesis studies whereas the displacement of the first repeat and the presence of nonnative interactions at the TSE are simulation results which supplement the experimental data. Furthermore, the unfolding trajectories reveal the presence of two on-pathway intermediates with partial α-helical structure. Finally, on the basis of the available experimental and simulation results we suggest that in modular proteins the shift of the folding TSE toward the native structure upon reduction of the number of tandem repeats is consistent with the Hammond effect. Proteins 2006;64:178–192.** © 2006 Wiley-Liss, Inc.

Key words: ankyrin repeat proteins; flexibility; Φ-value analysis; folding pathways; Hammond behavior; cancer

INTRODUCTION

Proteins consisting of a linear chain of homologous structural units, called repeats, are very common in nature and prevalently mediate protein–protein interactions. The ankyrin repeat consists of 33 amino acids forming a loop, a β-turn, and two antiparallel α-helices.[1] Several ankyrin repeats are clustered in a linear array to form an elongated structure that is stabilized predominantly by short-range interactions between residues close in sequence. Moreover, the hydrophobic core has a toroidal shape unlike in globular proteins.

$p16^{INK4a}$ (Fig. 1), also referred to as multiple tumor suppressor 1 (Ref. 2) and abbreviated as p16 in this report, is a well-studied four-repeat protein, especially because of its direct link between cancer and cell cycle control. p16 regulates the proliferation of the cell by binding to cyclin D1 dependent kinases CDK4 and CDK6, and thus inhibiting them. The latter proteins are responsible for the inactivation of the retinoblastoma protein (Rb) through phosphorylation. If not phosphorylated, Rb blocks transcription by inhibiting a specific transcription factor such that the cell cannot proliferate. These functional relations are known as the *p16/CDK4/CDK6/cycD1/Rb pathway*, where p16 and Rb are two tumor suppressors. In fact, they are found mutated in cancerous tissues (for a review, see Lukas et al.[3] and Serrano[4]).

Experimentally, p16 was found to be marginally stable[2,5] and with a tendency to aggregate, which is drastically increased by single point mutations.[2,6] This might explain its deactivation in cancerous tissues, where it is often found mutated. Inactivation of p16 is second only to loss of p53 function as the most frequent event observed in human tumors. The nuclear magnetic resonance (NMR) solution structure first published in Byeon et al.[7] and later refined in Yuan et al.[8] revealed that the first helix of the second ankyrin repeat consists of only one turn. Furthermore, compared with the helix bundles, the loops have a less-defined structure because of the conformational flexibility in these regions. Mutagenesis experiments have

The Supplementary Material referred to in this article can be found at http://www.interscience.wiley.com/jpages/0887-3585/suppmat/

Grant sponsors: National Competence Center in Research (NCCR) in Structural Biology; Swiss National Science Foundation.

Giovanni Settanni's present address is MRC Center for Protein Engineering, Cambridge, UK.

Correspondence to: Amedeo Caflisch, Department of Biochemistry, University of Zürich, 8057 Zürich, Switzerland. E-mail: caflisch@bioc.unizh.ch

Received 8 August 2005; Revised 9 November 2005; Accepted 8 December 2005

Published online 4 April 2006 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.20953

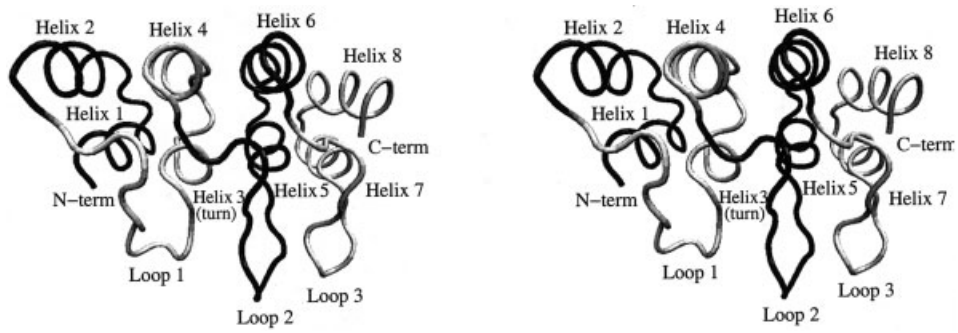

Fig. 1. Backbone of the NMR structure of p16[8] (stereoview). The repeats are distinguished by different gray shades and the N-terminal repeat is on the left.

shown that residues belonging to the second and third loop are functionally important for binding to CDK4.[9] The crystal structure of the complex p16/CDK6 reveals that the functionally important residues for binding to CDK6 are mostly located in the helical turn of the second repeat, but also in the tips of the first and third loop and in the helical regions of the first three repeats.[10] The third and fourth repeat have been identified as the minimum folding unit of p16.[11] Taken together, the experimental studies suggest that the two N-terminal repeats are functionally important whereas the two C-terminal repeats are mainly responsible for conformational stability.

The structure of p16 in its native state is crucial for its biological function. However, it is of interest to understand how it acquires its folded, biologically active conformation. Hence, it is useful to characterize all states during folding/unfolding: native, transition, intermediate, and denatured.[12] Urea induced unfolding/refolding experiments of p16 show the presence of an intermediate during refolding with less α-helical content than in the native state.[6] In contrast, unfolding seems to follow a two-state model.[6] Recently, mutagenesis experiments have been performed to evaluate a set of Φ values for the characterization of the transition state ensemble (TSE).[13] The Φ value is:

$$\Phi = \frac{\Delta G_{\ddagger - U} - \Delta G'_{\ddagger - U}}{\Delta G_{F-U} - \Delta G'_{F-U}} = \frac{\Delta\Delta G_{\ddagger - U}}{\Delta\Delta G_{F-U}} \qquad (1)$$

where $\Delta G_{\ddagger - U}$ and ΔG_{F-U} are the free energies of the transition and native states, respectively, relative to the unfolded state for the wild-type. A prime indicates the corresponding terms for the mutant. A value of Φ close to 1 indicates that the transition state and the native state are both destabilized by the same amount. In this case, the structure of the transition state at the site of the mutation is almost the same as in the folded state. Conversely, a Φ close to 0 means that in the transition state the region around the mutated residue is as unstructured as in the denatured state.[14] The Φ-value analysis of p16 indicates that in the transition state the two N-terminal repeats, especially the second ankyrin repeat, are partly unstructured whereas the C-terminal repeats are still conformationally close to the native state.[13]

All-atom simulation studies have been performed to characterize the unfolding/folding TSE of several proteins but not p16. Daggett and coworkers used structural analysis and conformational cluster analysis to identify the TSE of Chymotrypsin Inhibitor 2[15] and the on-pathway intermediate of the α-helical FF domain from human HYPA/FBP11.[16] Other studies have made use of the experimentally determined Φ values to isolate members of the TSE along molecular dynamics (MD) trajectories of acylphosphatase[17] and SH3 domains.[18,19] Recently, the TSE of a three-stranded antiparallel β-sheet peptide has been identified using the probability of folding (p_{fold})[20] and the TSE structures have been used for in silico Φ-value analysis.[21]

Here, the flexibility of p16 at physiologic temperature and its unfolding pathways at high temperature are investigated by all-atom MD simulations. The information content of experimental Φ values might not be sufficient to rigorously distinguish conformations belonging to the TSE from those not belonging to it.[22] Therefore, an approach based on both, experimental Φ values and sudden conformational changes along the unfolding trajectories, is used to identify TSE conformations. The present work was motivated by the following questions: Is there a relationship between the structural flexibility and binding properties of the different p16 ankyrin repeats? How broad is the TSE of p16 and which of its repeats contribute most to the conformational heterogeneity of the TSE? Is it possible to identify unfolding intermediates and do they have some of the repeats folded? The simulation results interpreted with the help of the known experimental Φ values provide a detailed picture of the unfolding process.

MATERIALS AND METHODS

Simulations

The MD simulations were performed with the program NAMD[23] using the CHARMM all hydrogen force field (PARAM22).[24] The runs were started from the first two NMR conformations in the Protein Data Bank (PDB) entry 1A5E and the last NMR structure from the PDB entry 1DC2 after having cleaved residues 1–13 and 135–156 because they were found unstructured by NMR spectroscopy. Acetyl and N-methylamide were used to cap the N-terminus and C-terminus, respectively. Two 100-ns unfolding simulations at 400 K were started from each initial conformation and they are called in this report US1 and US2 (from the first NMR conformation in 1A5E), US3

and US4 (from the second NMR conformation in 1A5E), US5 and US6 (from the last NMR conformation in 1DC2). When starting from the same structure, the seed was changed to generate different initial velocity distributions. Furthermore, a control simulation of 50 ns was performed at 300 K starting from the last NMR conformer of 1DC2. The protein structures were inserted into a cubic box of 17,576 preequilibrated water molecules with a side length of 80.6 Å. Forty-eight chloride and 49 sodium ions were added to neutralize the system and approximate 150 mM salt concentration. The water molecules overlapping with the protein or the ions were removed if the distance between the water oxygen and any atom of the protein or any ion was smaller than 3.1 Å. The simulated systems contained in total between 51,000 and 52,000 atoms. To avoid finite size effects, periodic boundary conditions were applied by replicating the waterbox at infinity in all three space directions. Electrostatics interactions were calculated directly within a cutoff of 12 Å, whereas long-range electrostatics effects were taken into account by the Particle Mesh Ewald summation method.[25] van der Waals interactions were treated with the use of a switch function starting at 10 Å and turning off at 12 Å. The temperature was kept constant by using the Berendsen thermostat with a relaxation time of 0.2 ps while the pressure was held constant at 1 atm by applying a pressure piston.

Before production, harmonic constraints were applied to the positions of all the atoms of the protein to equilibrate the system. The equilibration phase lasted 0.2 ns. The dynamics was integrated with a time step of 2 fs. The covalent bonds involving hydrogens were rigidly constrained by means of the SHAKE algorithm with a tolerance of 10^{-8}. Snapshots were saved every 2 ps for a total of 50,000 structures for each 100-ns MD run.

Nuclear Overhauser Effect (NOE) Violations and Native Contacts

NOE violations were calculated along the room temperature simulation. A NOE distance was considered violated, if the equation

$$r_{\min}^{\exp} \leq \langle r(t)^{-6} \rangle^{-1/6} \leq r_{\max}^{\exp} \tag{2}$$

was not fulfilled, where $r(t)$ is the interproton distance at simulation time t, r_{min}^{exp} and r_{max}^{exp} are the lower and upper distance limit, respectively, and $\langle\rangle$ represents a time average.

The conformations sampled at room temperature were also used to determine native hydrogen bonds and C_α native contacts. To define a hydrogen bond a H $\cdots$ O distance cutoff of 2.7 Å and a D—H $\cdots$ O angle cutoff of 120° were used. A C_α contact involves two C_α atoms with a distance smaller than 6.5 Å and belonging to residues that are separated in sequence by at least two other residues (i.e., pairs i, j with $j > i + 2$). Only those hydrogen bonds and C_α contacts present in at least two-thirds of the simulation frames were selected as native contacts. They were used to describe the conformational flexibility of p16 at room temperature and the changes in secondary and tertiary structure during unfolding.

Identification of the TSE for Unfolding

A correlation has been observed between experimentally determined Φ values [Eq. (1)] and the fraction of hydrophobic side-chain contacts formed in the transition state relative to the native state.[26] Hence, the calculated Φ value for a residue i, Φ_i^{calc}, in an MD simulation can be approximated by[15]:

$$\Phi_i^{\text{calc}} = \frac{N_i^{\text{Conf}}}{N_i^{\text{F}}} \tag{3}$$

where N_i^{Conf} is the number of native side-chain contacts of residue i in a given conformation, and N_i^F is the average number of native side-chain contacts during the run at 300 K. Native side-chain contacts are those persistent during at least two-thirds of the room temperature simulation. Side-chain contacts are considered to exist when any side-chain heavy atom of a residue i is closer than 6 Å to heavy atoms of nonadjacent residues (residues i and $i \pm 1$ are omitted).

To monitor the average distance from the experimentally determined TSE, the root-mean-square deviation (RMSD) between the experimentally determined Φ values (Φ_i^{exp}) and Φ_i^{calc} was evaluated along the unfolding trajectories at 400 K[17,18]:

$$\Phi\text{-RMSD} = \sqrt{\frac{1}{M_\Phi}\sum_i [\Phi_i^{\text{calc}} - \Phi_i^{\text{exp}}]^2} \tag{4}$$

where M_Φ is the number of selected Φ_i^{exp} values ($M_\Phi = 11$; Ala^{30}, Ala^{36}, Val^{51}, Met^{54}, Ala^{60}, Ala^{86}, Val^{95}, Val^{115}, Val^{126}, Ala^{127}, Ala^{133}). Ser^{43} was excluded because it has a $\Phi^{\exp}$ value of -0.36 and no side-chain contacts in the 300 K simulation. The five residues in the second (Ala^{73}, Ala^{76}) and third loop (Ala^{102}, Val^{106}, Ala^{109}) as well as the C-terminal residue Ala^{134} were omitted, because their Φ_i^{calc} presented large fluctuations (in the range between 0 and 1). Moreover, the mutants Ala76Gly and Ala134Gly have a stability change smaller than 0.5 kcal/mol which results in high uncertainty of $\Phi^{\exp}$.[21,27,28] Because of large fluctuations, the running average of the Φ-RMSD with a time window of 0.4 ns was plotted for each simulation. The latter presented several local minima with values lying mostly between 0.15 and 0.2. To identify which one of the local minima corresponds to the transition state, a more detailed analysis of the structural changes during the simulation was necessary.

The structure of a protein at the transition state is expected to change rapidly. Thus, in addition to the Φ-RMSD, the rate of structural changes[15] along each unfolding trajectory has been monitored to locate the transition state. The structural changes have been characterized by the C_α RMSD of a structure with respect to the preceding structure. The resulting time series gives information about the location in time of sudden conformational changes. Because there are large fluctuations, the analysis was done on a running average of the RMSD with a time window of 2 ns. A rapid conformational change is defined to occur when the RMSD from the preceding

structure exceeds the mean value during the simulation. The transition state was searched in the vicinity of such sudden conformational changes, more precisely in the set of snapshots contained in a time window ranging from the local minimum to the local maximum of the RMSD immediately preceding respectively following the time point when it exceeds the mean value. During an unfolding process there might be several energy barriers to be passed, after which the protein can undergo a rapid conformational change. For this reason, several sets of conformations satisfy these criteria along the unfolding trajectories (Fig. 2). However, the transition state of p16 unfolding should occur early in the unfolding trajectory[13]; this hypothesis is validated later by calculating the β_T value from the solvent accessible surface area (SASA) of the identified TSE conformations. Hence, the first set of conformations that fulfilled both criteria (i.e., it preceded a rapid conformational change and its time averaged Φ-RMSD was not larger than 0.2) was considered. Twenty conformations lying on a time interval of 0.4 ns centered at the local minimum of the Φ-RMSD were used to describe the transition state of a simulation. Although the TSE conformations were identified using running averages of the Φ-RMSD and of the C_α RMSD, the coordinates of the TSE snapshots were never averaged over time to avoid artificial conformations with physically unrealistic atomic positions. Figure 2 illustrates the approach used to identify the TSE.

One validation of the identified TSE is the computation of the β_T value.[29] It was previously shown that the SASA correlates to kinetic and equilibrium *m*-values, i.e., the location of the TSE on the unfolding reaction coordinate.[30] Assuming a general proportionality between *m*-values and changes in SASA, β_T can be estimated as[30]:

$$\beta_T = \frac{\mathrm{SASA_U} - \mathrm{SASA_{TSE}}}{\mathrm{SASA_U} - \mathrm{SASA_F}}. \quad (5)$$

The SASA of the unfolded state was estimated as the largest value of solvent exposed surface found in the simulations, whereas the SASA of the native state was averaged over all 20 NMR conformations. Assuming a two-state model, β_T can be derived from the experimentally determined equilibrium *m*-values:

$$\beta_T = 1 - \frac{m_{\mathrm{off}}}{m_{\mathrm{on}} + m_{\mathrm{off}}}. \quad (6)$$

In previous works by others and us,[18,22] the probability of folding before unfolding, p_{fold}, was used to validate the identified TSE conformations. A p_{fold} close to 0.5 validates the identified TSE, because in a simulation started from a transition state conformation the protein is expected to fold before unfolding in 50% of the cases. However, because the determination of p_{fold} needs a large number of simulations to have statistical significance, it is only feasible for systems that require a relatively low computational cost, such as implicit solvent[18] or Monte Carlo Go model[22] simulations. Because the system studied here contained an explicit treatment of the solvent, the computational requirements would have been prohibitive. However, the TSE identified in the simulations is validated quantitatively by comparison with the experimental Φ values not used in the evaluation of Φ-RMSD and *m*-value analysis.

Conformational Cluster Analysis for the Identification of Intermediates

An approach to identify low free-energy intermediate states in MD simulations was proposed by Jemth et al.[16] If a state has the probability p of being observed in one simulation, it will have the probability p^2, p^3, and so on of being observed in two, three, etc., simulations. For high energy states it can be assumed that $p \ll 1$, thus their probability of being observed in multiple simulations decreases exponentially with the number of MD runs. Figure 3 shows the C_α RMSD of the first two repeats between all structures sampled during the six simulations. A similar plot with the C_α RMSD of the whole protein (not shown) is qualitatively similar but presents more noise. The on-diagonal clusters (consisting of conformations with a C_α RMSD with respect to each other smaller than 6 Å) represent regions of local similarity during unfolding whereas the off-diagonal clusters (because of conformations with a C_α RMSD with respect to each other smaller than 8 Å) indicate that similar conformations are being sampled in different simulations. Although equilibrium is not reached during the simulated time, partial refolding events in two simulations indicate that a large amount of conformational space has been explored allowing the description of the relevant steps during unfolding. The C_α RMSD matrix alone is not sufficient to classify the on-diagonal clusters into intermediates, because conformations in different simulations with an RMSD with respect to each other larger than 6 Å still presented similar secondary structure elements in the two N-terminal repeats. Thus, the on-diagonal clusters were grouped in two classes of partially unfolded intermediate states based on their native α-helical content (measured using the C_α RMSD from the NMR structure). In all simulations, the cluster that followed in time the transition state presented intact α-helices in the first repeat, in fact their C_α RMSD from the native NMR structure was never larger than 3 Å. These clusters were called intermediate I_1. In most simulations, the conformations in the clusters sampled after intermediate I_1 presented a partly or completely unfolded first repeat, in fact the C_α RMSD from the native state of at least one of the helices was larger than 3 Å. Those clusters were called intermediate I_2. In the simulation US2, a refolding of the first repeat was observed in a cluster that followed in time intermediate I_2. This cluster was also classified as belonging to intermediate I_1. Both intermediates are indicated in Figure 3 next to the corresponding clusters.

RESULTS

Flexibility at 300 K

The time series of the C_α RMSD (Fig. 4) and the plot of C_α atom fluctuations (Fig. 5) at 300 K show that the loops are more flexible than the helix bundles. Most impor-

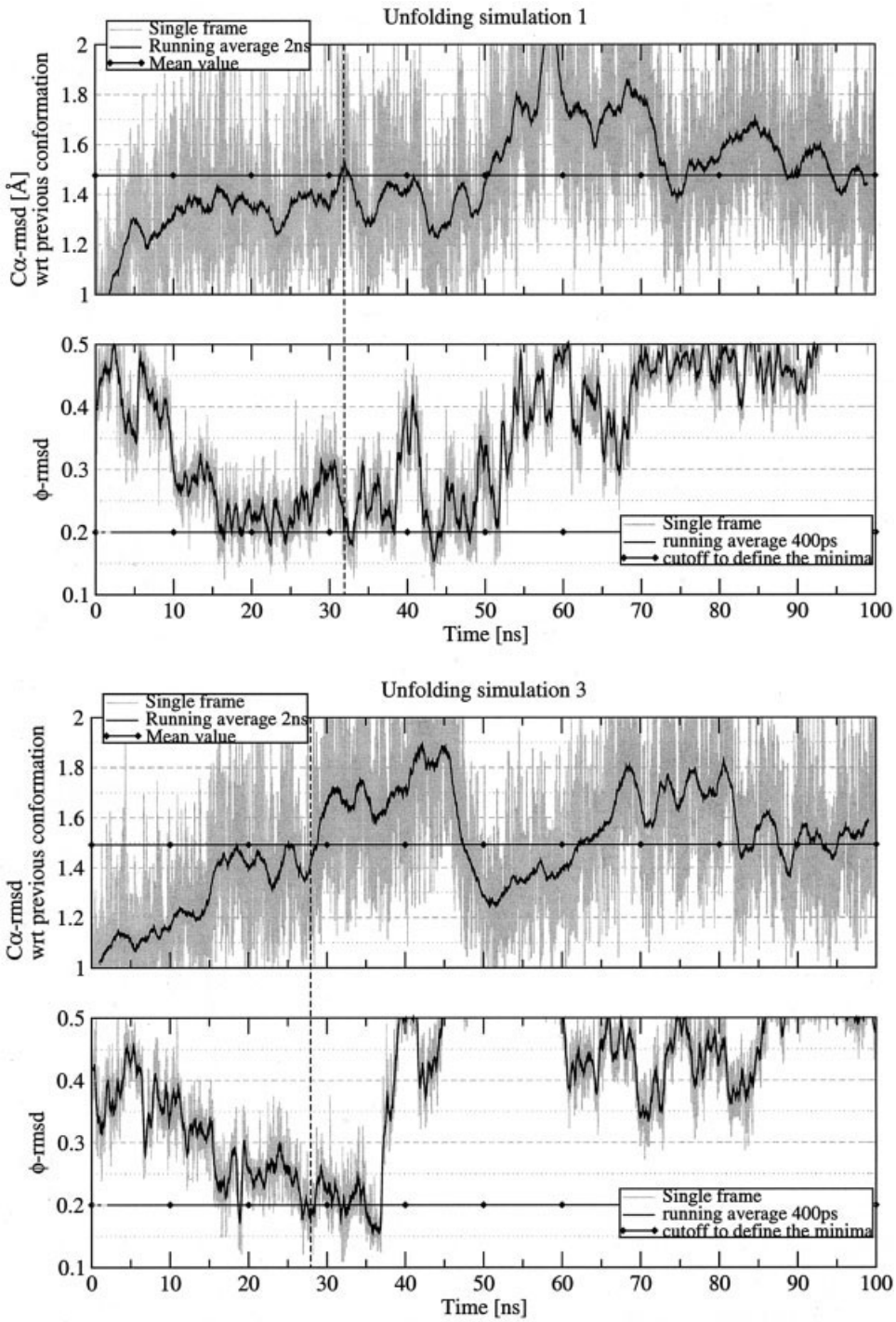

Fig. 2. Identification of members of the TSE along two trajectories at 400 K. Time series of the running average over 2 ns of the RMSD of the C_α atoms with respect to the previous conformation in the trajectory (it gives information about rapid conformational changes) and the running average over 400 ps of the Φ-RMSD. The conformations preceding a rapid conformational change and presenting a low Φ-RMSD are identified as members of the TSE. The vertical dashed line marks the time point in the simulation when the TSE is reached. The first 0.2 ns of equilibration are not shown.

tantly, repeats 2 and 4 show the largest fluctuations which are mainly the result of loop motion observed in the MD trajectory and NMR analysis except for the structural plasticity of the loop of the second repeat which is much higher in the MD simulation than in the 20 NMR conformers (Fig. 5). This discrepancy could be a consequence of the force field and simulation protocol, insufficient NMR data, or both. Interestingly, the largest deviation between the

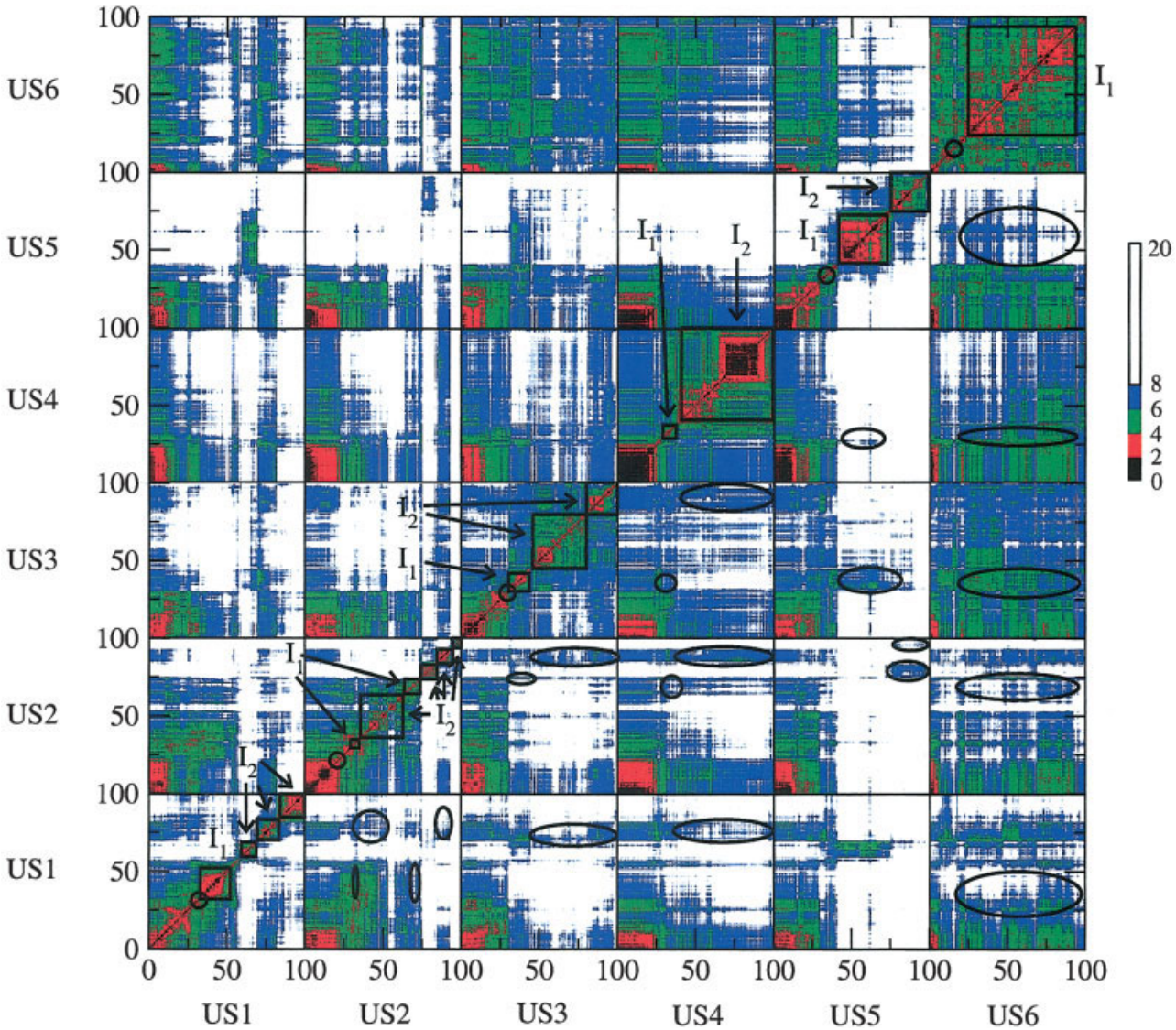

Fig. 3. Matrix of the values of the RMSD of the C_α atoms in the first two repeats. Each conformation saved along the six unfolding simulations is compared with all other conformations.[16] The *x* and *y* coordinates represent the time points in the different simulations. The values for the RMSDs are color-coded and the legend for the magnitudes is given in the color scale on the right of the plot. On-diagonal clusters (boxed off) indicate local clusters. Off-diagonal clusters (circled) suggest that the on-diagonal clusters belong to a common unfolding intermediate. Structural analysis suggested the grouping of the on-diagonal clusters into two distinct unfolding intermediates as indicated by the labels I_1 and I_2. The time points when the transition state was reached in the different simulations are marked with a circle on the diagonal.

solution structure of free p16 and the X-ray structure of its complex with CDK6 is located in the loop of the second repeat (see Fig. 9 of Ref. 8), but it is not clear if this deviation is a consequence of crystal packing or a conformational change associated with CDK6 binding.

Excluding the loops, most of the RMSD is accounted for by the rigid body motion of the individual repeats with respect to each other (Fig. 4). Visualization shows a tilting of about 10° of the first and last repeat relative to the central repeats which is consistent with the displacement observed in the crystal structure of a ribonuclease inhibitor, a leucine-rich repeat protein, upon binding to ribonuclease.[31] Hence, the rigid body motion observed in p16 could be functionally relevant.

On average, 11% and 24% of the total number of NOEs and long distance NOEs, respectively, are violated along the trajectory indicating that the protein is conformationally flexible at room temperature. About 20% of the NOE distances involving residues in the loops are violated, whereas in the helix-bundles only 9% are violated. This is consistent with the fact that 71% of the measured NOEs involve the helix bundles. In fact, the latter were found in NMR experiments more rigid than the loops.[5] The native contacts listed in Table I give more detailed information about the conformational flexibility of the repeats relative to each other. Repeat 2 presents a smaller number of native contacts than the other three repeats in agreement with its reduced amount of secondary structure (its first

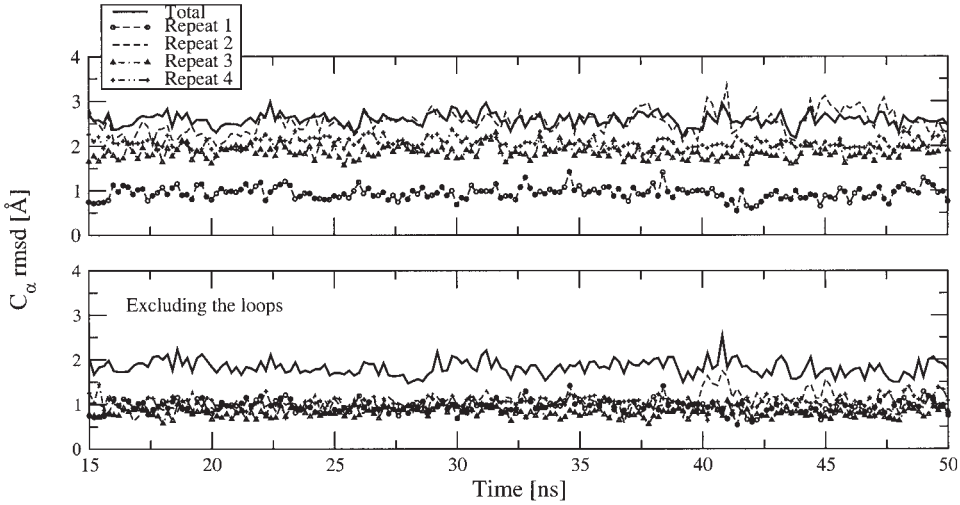

Fig. 4. Time series of the C_α RMSD with respect to the NMR conformation in the simulation at 300 K. The solid thick line without symbols is used for the whole protein, the thin lines for individual repeats. The first 15 ns of the simulation are considered equilibration and are not shown. **Top:** C_α RMSD including all secondary structural elements and the loops. The second repeat accounts for most of the total C_α RMSD. **Bottom:** C_α RMSD including only the helix-turn-helix motifs. The separation of the solid thick line from the other lines is the result of the rigid-body motion of the individual repeats with respect to each other.

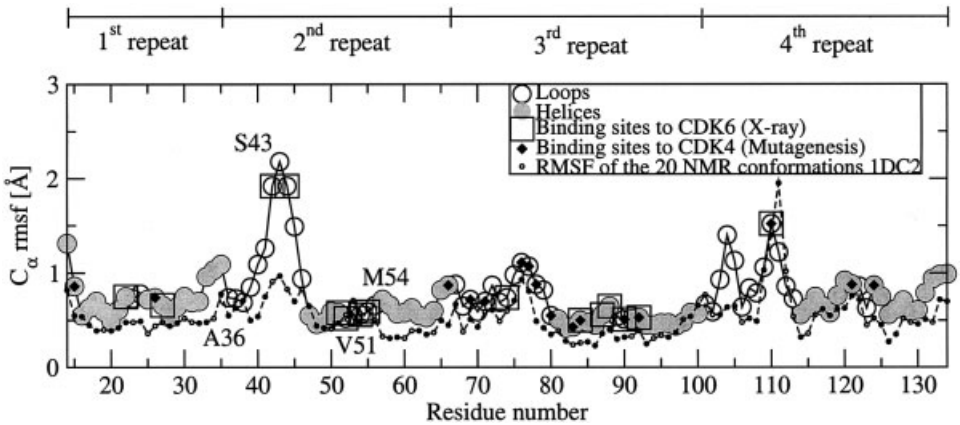

Fig. 5. Flexibility of p16 at 300 K. The data show the RMS fluctuations (RMSF) of the C_α atoms of the snapshots sampled during the simulation (large circles) and of the 20 NMR conformers (small circles). The conformations were fitted to the average structure during the simulation and the average NMR conformation, respectively. Residues with experimental Φ-values close to 0[13] are indicated by labels. The p16 residues interacting with CDK4 as detected by mutagenesis[9] and the residues in contact with CDK6 in the X-ray structure[10] are marked by diamonds and squares, respectively. It can be observed that some but not all of the very flexible residues are involved in binding.

TABLE I. Native Contacts at 300 K†

	Total	R1	R2	R3	R4	R1–R2	R2–R3	R3–R4
Hydrogen bonds	73	17	12	18	17	1	4	4
C_α contacts	165	34	25	32	39	11	11	13

†R1 to R4 denote the intrarepeat contacts in repeat 1 to repeat 4 whereas R1–R2, R2–R3, and R3–R4 denote the interrepeat contacts. Their computation is described in the Materials and Methods section.

helix consists of only one helical turn) compared with the canonical ankyrin fold. This scarce presence of secondary structural elements could be one reason for the relatively large C_α RMS fluctuations in the first part of the second repeat (Fig. 5). However, it is difficult to correlate the local flexibility of p16 at room temperature and the binding functionality of each repeat. Namely, not only the conformationally flexible tips of the loops but also the relatively more rigid helix bundles have an important functional role in binding to CDK4 and CDK6 (Fig. 5).

TABLE II. Properties of the TSE Conformations[†]

TSE	Time (ns)[a]	ρ[b]	SASA (Å^2)[c]	β_T	H-bonds[d]	C_α contacts[e]
US1	31.87	0.64 (0.97)	7,370	0.77	0.62	0.74
US2	30.05	0.78 (0.96)	7,310	0.79	0.55	0.71
US3	27.73	0.71 (0.95)	7,407	0.76	0.57	0.70
US5	28.07	0.73 (0.94)	7,656	0.68	0.54	0.70
US6	13.63	0.77 (0.95)	7,097	0.86	0.54	0.72

[†]All values are averaged over 400 ps.
[a]Time point in the simulation when the transition state is reached.
[b]Pearson correlation coefficient between 18 (11 less fluctuating) Φ_i^{calc} and Φ_i^{exp} values.
[c]SASA of all residues used to estimate β_T. The value for β_T obtained from the experimentally determined equilibrium *m*-value is 0.84.
[d]Fraction of formed native hydrogen bonds.
[e]Fraction of formed native C_α contacts.

TABLE III. Averaged Properties of the TSE Conformations

	Total	R1	R2	R3	R4	R1–R2	R2–R3	R3–R4
H-bonds[a]	0.56	0.54	0.43	0.61	0.81	0	0.03	0.49
C_α contacts[a]	0.71	0.80	0.60	0.82	0.91	0.11	0.36	0.65
SASA hydrophobic[b]	1.21	1.16	1.70	1.03	1.04			

[a]Fraction of native hydrogen bonds (H-bonds) and C_α contacts.
[b]SASA of hydrophobic residues relative to the corresponding averaged values of the simulation at 300 K. R1 to R4 denote repeat 1 to repeat 4 whereas R1–R2, R2–R3, and R3–R4 denote the interrepeat contacts.

Unfolding TSE

The TSE was identified along five of the six trajectories by monitoring both the deviation from the experimental Φ values (Φ-RMSD) and large conformational changes (Fig. 2 and Materials and Methods). In the unfolding simulation US4, the running average of the Φ-RMSD never reached a value smaller than 0.25 because the two C-terminal repeats lose their tertiary structure concomitantly with the unfolding of the two N-terminal repeats. The behavior of US4 might indicate the presence of multiple unfolding routes but the available statistics (six MD runs) do not allow us to draw stronger conclusions on multiple pathways. Interestingly, two main folding pathways have been very recently observed in MD simulations of a C_α-based Go model of p16.[32] Moreover, multiple folding/unfolding routes have been suggested for the *Drosophila* Notch receptor[33] on the basis of equilibrium denaturation studies of several constructs containing subsets of the seven ankyrin repeats.

The properties of the TSE determined in the five simulations are summarized in Tables II and III. The TSE conformations can be validated by the β_T (β Tanford) value,[29] which is a measure of how much surface is solvent exposed in the TSE relative to the native state (see Materials and Methods section). The β_T value calculated from the SASA of the isolated TSE conformations [Eq. (5) in Materials and Methods section] is on average 0.77. This value is close to the β_T of 0.84, calculated from the experimentally determined *m*-values[13] [Eq. (6) in Materials and Methods section]. This relatively large value of β_T suggests that the TSE of p16 is very compact and it is positioned close to the native state on the folding reaction coordinate. It also justifies our assumption to identify those conformations that are sampled early in the unfolding simulation as members of the TSE (see Materials and Methods section).

Further validations of the identified TSE conformations are the correlations between Φ_i^{calc} and Φ_i^{exp}. The Pearson linear correlation coefficient (ρ) between the 11 Φ_i^{exp} values and those obtained by averaging Φ_i^{calc} over the identified TSE conformations of five unfolding simulations is 0.98 [Fig. 6(a)]. A ρ value of 0.73 is obtained by considering 17 of the 18 Φ_i^{exp} values [Fig. 6(b)]. Ser^{43} is neglected because it does not have side-chain contacts in the room temperature simulation which makes impossible the determination of its Φ_i^{calc} value. Moreover, its Φ_i^{exp} value deviates the most from the 0–1 range among the 18 residues ($\Phi_{Ser^{43}}^{exp} = -0.36$). A cross-validated correlation coefficient of 0.65 is measured on the six residues (Ala^{73}, Ala^{76}, Ala^{102}, Val^{106}, Ala^{109}, and Ala^{134}) not used to isolate the TSE conformations. This result is likely to originate from the fact that the 11 Φ_i^{exp} values used to identify the TSE are spread over the whole sequence and span all of the four repeats.

The total C_α RMSD between the different transition state conformations is quite large ranging from 4.2 to 11.9 Å and 2.2 to 9.4 Å for the whole protein and excluding the loops, respectively. (For the computation of the minimal value of the C_α RMSD, only snapshots belonging to different simulations were compared, i.e., fully uncorrelated conformations.) However, the diversity among the structures arises from the N-terminal half and mainly the second repeat [Fig. 6(c)]. Namely, the C_α RMSD between the sampled TSE conformations including only the two C-terminal repeats and the second helix of the second repeat ranged from 1.69 to 4.3 Å and 0.74 to 2.47 Å for all residues and excluding the loops, respectively. These relatively low values indicate that the sampled TSE conformations are similar to each other in the C-terminal half of p16 which is consistent with the available experimental data. However, the conformational diversity in the

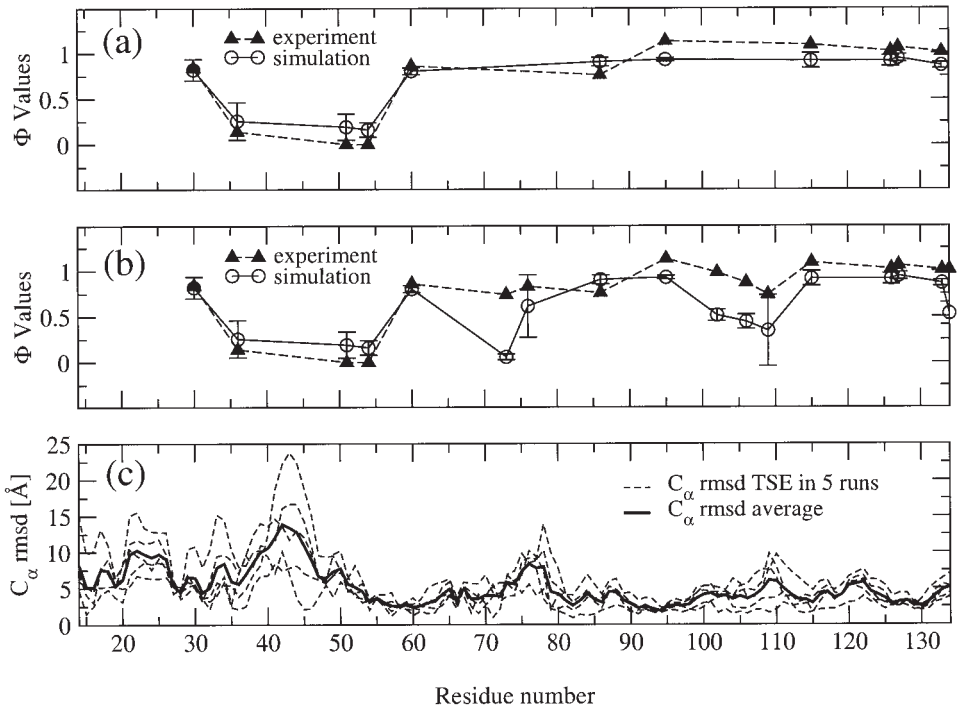

Fig. 6. Properties of the TSE. (**a**) and (**b**): Comparison of Φ_i^{calc} with Φ_i^{exp}. (**a**) The 11 Φ_i^{exp} used to identify the TSE by the Φ-RMSD criterion. (**b**) The 17 residues with well-defined Φ_i^{exp}. (**c**) C_α RMSD from the NMR structure as a function of residue number. The five dashed lines are averages over the TSE structures sampled in individual unfolding simulations.

two N-terminal repeats, which is consistent with the interpretation of the experimental Φ-value analysis, indicates that a significant amount of conformational space in the vicinity of the transition state was sampled in the five unfolding simulations.

The number of conserved native contacts was used to further characterize the flexibility of the two N-terminal repeats in the TSE. In this region on average only 50% of native hydrogen bonds and 62% of native C_α contacts are formed in the identified TSE conformations. Hence, this part of the protein is less structured in the TSE than the two C-terminal repeats, where 68% of the native hydrogen bonds and 83% of the native C_α contacts are present. For this reason, the first two N-terminal repeats in the identified TSE conformations were analyzed in further detail. It was found that whereas the secondary structure of the fourth helix is still preserved, the third short helix is completely unstructured in three of five runs (Fig. 7). Relative to the native state, the second repeat presents a larger solvent exposed surface of hydrophobic residues than the remainder of the protein, especially the region containing the loop and the helical turn (Fig. 8). A major contribution to the increase of SASA in the second repeat arises from three proline side-chains, Pro^{38}, Pro^{41}, and Pro^{48}, which in the TSE present on average a SASA of 70 ± 40, 95 ± 35, and 51 ± 32 Å^2, respectively, compared with a SASA in the 300 K run of 5 ± 6, 12 ± 8, and 0 ± 1 Å^2, respectively (Fig. 9). The interrepeat interactions between the second repeat and the adjacent repeats are mostly disrupted (only 29% of the interrepeat C_α contacts and 12% of the interrepeat hydrogen bonds are formed).

Nonnative Interactions at the TSE

Visualization of the identified TSE structures shows that the first loop is distorted compared with the native state (Fig. 10). In contrast, the first repeat retains its overall fold (54% and 88% of the native hydrogen bonds and native C_α contacts, respectively, are formed). However, the first repeat appears to undergo a rigid-body rotation of up to 90° in both directions with respect to the other repeats and a translation of up to 3 Å parallel to the helix bundles (only 16% of the C_α contacts and none of the hydrogen bonds between the first and second repeat are formed). In most TSE conformations, the first helix is twisted by about 90° clockwise when viewed from the N-terminus. As a consequence, the side-chain of Trp^{15} is partially buried at the interface between the first two repeats with a lower SASA, i.e., 71 ± 59 Å^2 (averaged over the TSE conformations of US2, US3, US5, and US6), relative to the 300 K run, where its SASA measures 177 ± 22 Å^2 (Fig. 9). Moreover, the Trp^{15} side-chain is involved in a larger amount of side-chain contacts in the TSE, i.e., 132 ± 50 (averaged over the TSE conformations of all unfolding runs except US1), than in the 300 K simulation where it presents on average only 57 ± 29 side-chain contacts. Furthermore, two hydrophobic amino acids in the second loop, Pro^{75} and Ala^{76}, as well as Leu^{104}, are better shielded from the solvent in the TSE than in the 300

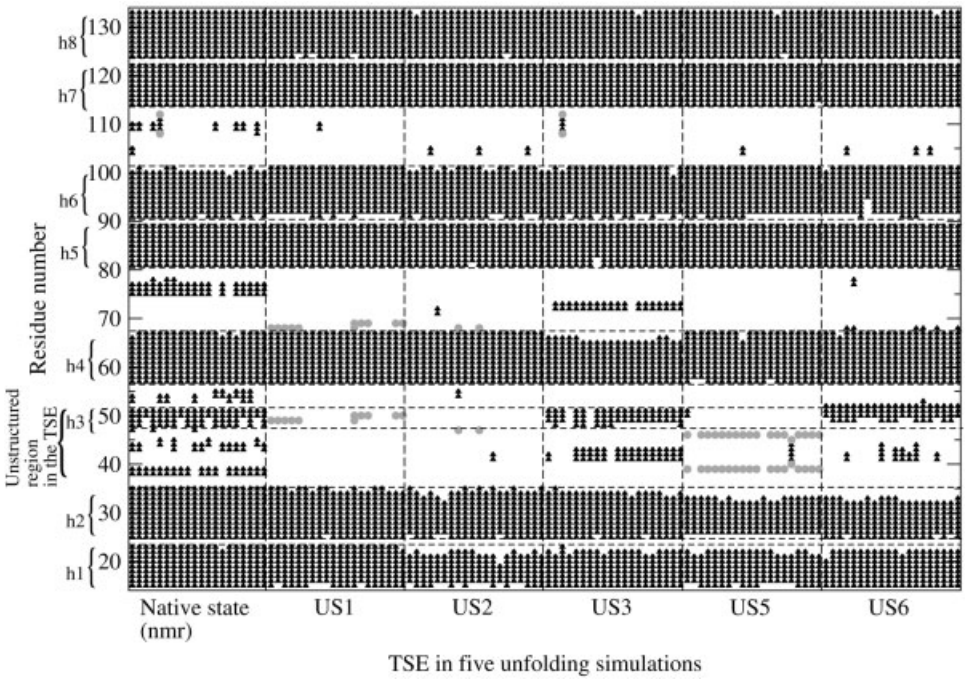

Fig. 7. Secondary structural elements of the 20 NMR conformers (PDB entry 1DC2) and of the 100 TSE conformations. The helices and the region found unstructured in the TSE are labeled on the left side of the plot. Black triangles and gray circles refer to amino acids involved in helical (α, π, and turn) and β structure (β-bridge and β-ladder), respectively, computed with DSSP.[42]

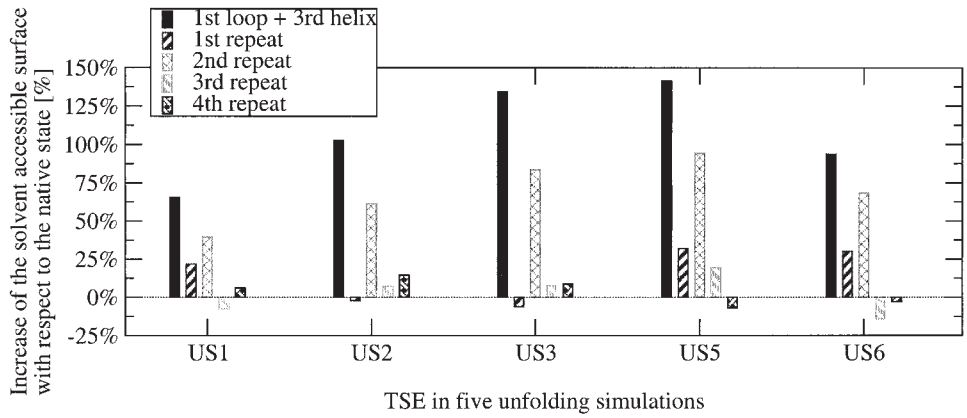

Fig. 8. Percentage of the SASA of hydrophobic residues relative to the averaged values of the simulation at 300 K. In each of the five unfolding simulations, where the TSE could be identified, the SASA was averaged over conformations sampled during 400 ps.

K run (Fig. 9), and are involved in nonnative interactions, as in the case of Trp^{15}. In fact, the SASA of the side-chains of Pro^{75}, Ala^{76}, and Leu^{104}, averaged over all conformations of the TSE, is 33 $\pm$ 29, 11 $\pm$ 14, and 21 $\pm$ 42 Å^2, respectively, which is smaller than in the 300 K run 123 $\pm$ 5, 55 $\pm$ 16, and 104 $\pm$ 38 Å^2, respectively. Residues Pro^{75}, Ala^{76}, and Leu^{104} also present in the TSE conformations a larger number of side-chain contacts, i.e., 55 $\pm$ 21, 17 $\pm$ 6, and 81 $\pm$ 29, respectively, than in the 300 K run, i.e., 8 $\pm$ 3, 5 $\pm$ 2, and 23 $\pm$ 20, respectively. Visual analysis shows that Pro^{75} is buried at the interface between the third and fourth repeat in the TSE conformations of all unfolding runs, except US6 where it is located at the interface between repeat 2 and 3. The Ala^{76} side-chain is sequestered at the interface between the second and third repeat in all TSE conformations. The Leu^{104} side-chain is buried at the interface between third and fourth repeat in most TSE conformations and in the 20 NMR conformations, but it is exposed to solvent in the 300 K run. These simulation results provide further evidence for the presence of nonnative interactions at the TSE.

Intermediates in the Unfolding Pathway

Two partially folded intermediate states were observed in all MD trajectories (Fig. 10). During denaturation, the intermediate I_1 is found on-pathway between the TSE and

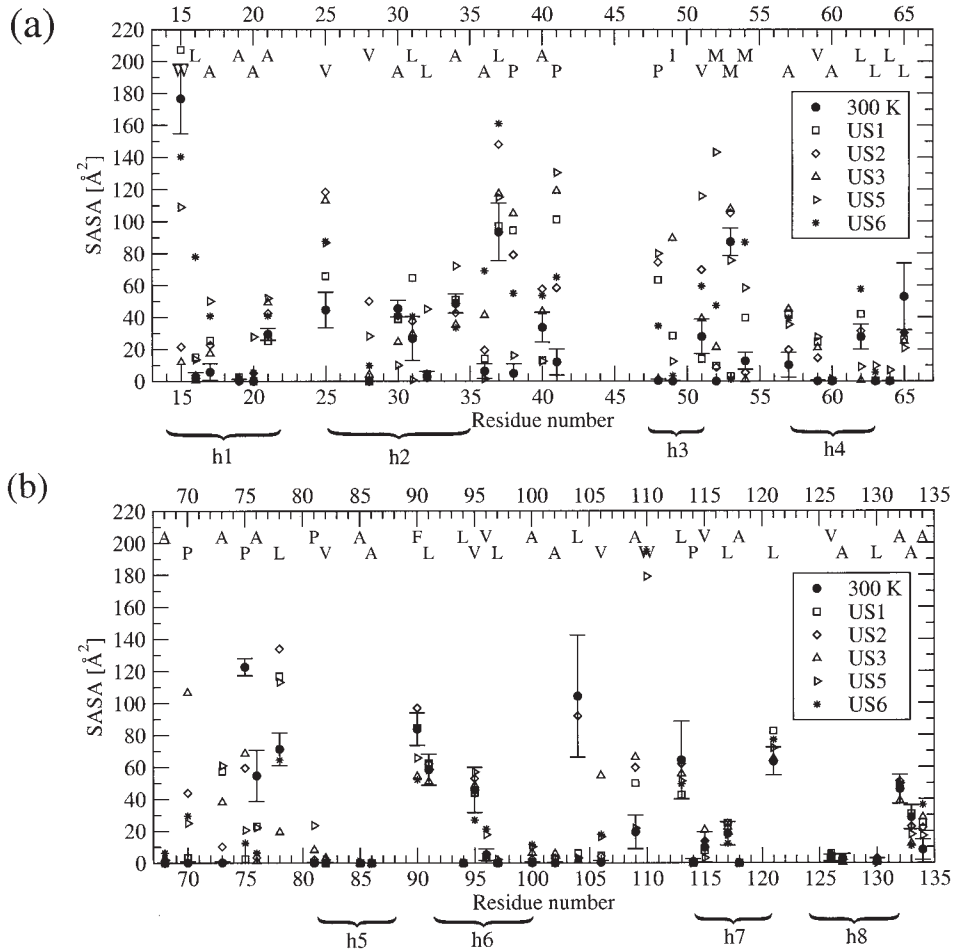

Fig. 9. SASA of hydrophobic side-chains in the TSE. (**a**) N-terminal repeats 1 and 2. (**b**) C-terminal repeats 3 and 4. The standard deviation is indicated by error bars only for the values corresponding to the 300 K simulation to avoid overcrowding of the plot. The hydrophobic residues are labeled by one letter code at the top of the plot. The secondary structure elements of the native state are labeled at the bottom of the figure.

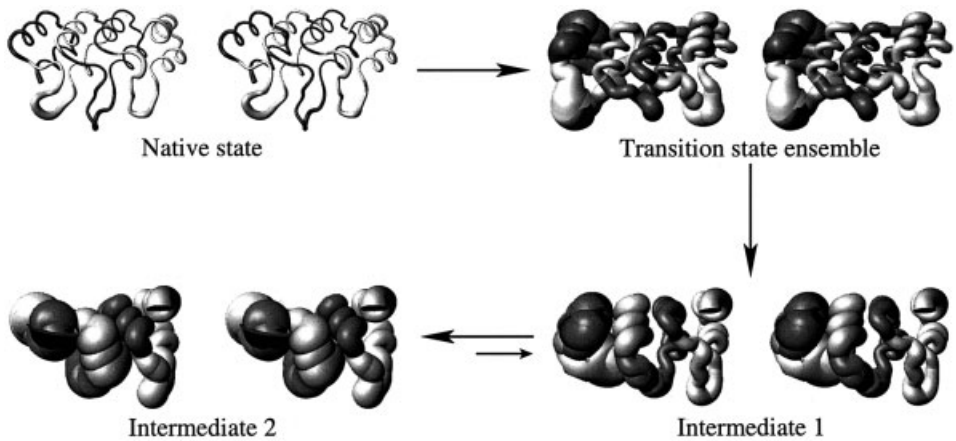

Fig. 10. The unfolding pathway of p16 (stereoview). The thickness of the tubes represents the RMS fluctuations of the C_α atoms. The double arrow between both intermediates indicates that in a simulation the protein switched back to intermediate 1. The figure was made with MOLMOL.[43]

the intermediate I_2. The population of I_1 and I_2 varies substantially (Fig. 3). The first loop and the first helix of the second repeat are unstructured in I_1 [Figs. 10 and 11(a)]. About 76% of the native contacts are conserved in repeat 1 [Table IV(a)] which is displaced with respect to its position in the native state (Fig. 10). The slightly lower

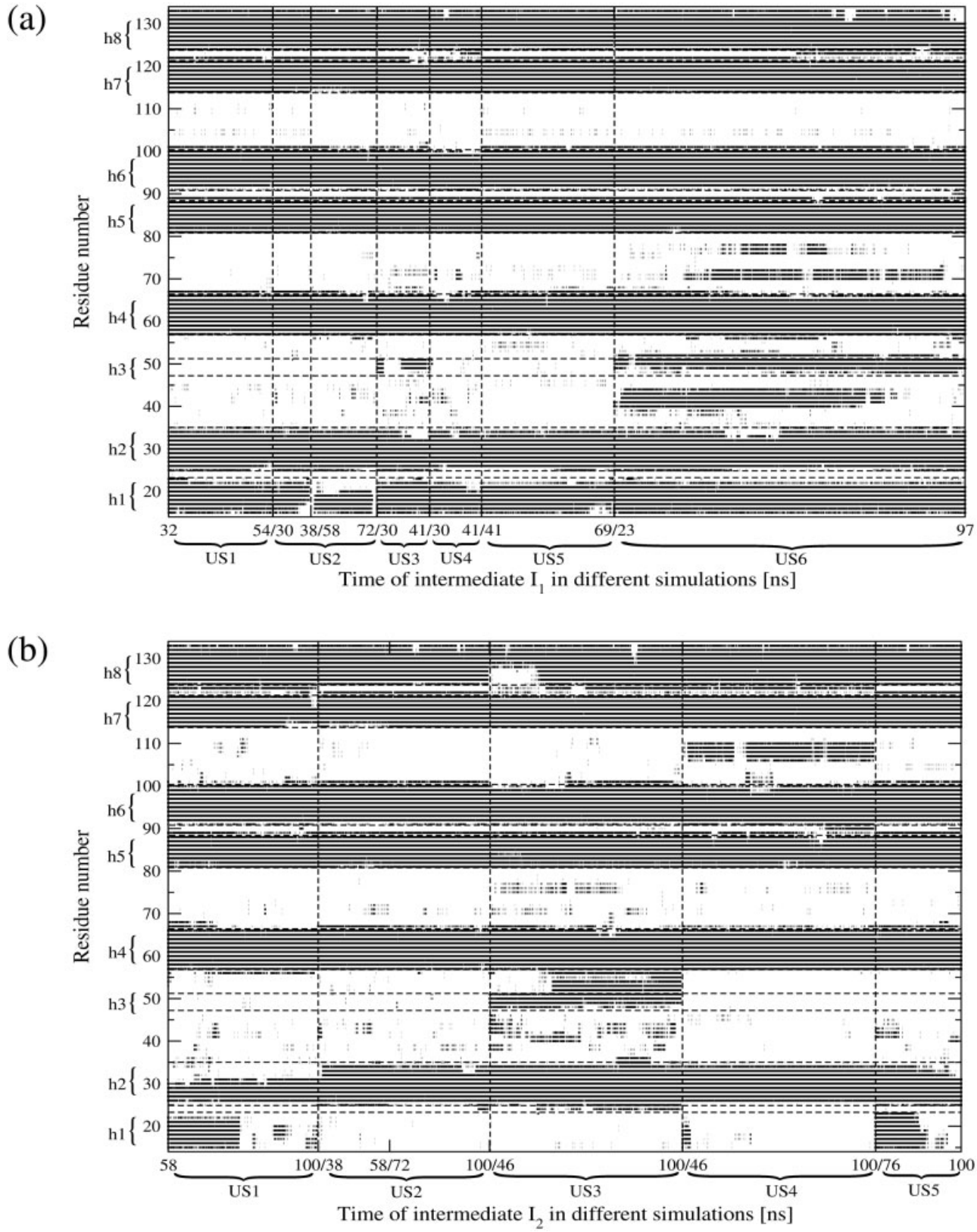

Fig. 11. Helical structure content in the conformations belonging to intermediate 1 (**a**) and 2 (**b**) calculated with DSSP.[42] The helices are labeled on the left side of the plots. The start and end times of the clusters of conformations belonging to the intermediate 1 or 2 for the six unfolding simulations are reported on the *x*-axes.

TABLE IV. Averaged Properties of the Intermediates

	Total	R1	R2	R3	R4
Intermediate 1					
H-bonds[a]	0.53	0.51	0.38	0.60	0.70
C_α contacts[a]	0.67	0.76	0.62	0.80	0.84
SASA hydrophobic[b]	1.20	1.15	1.54	1.10	1.08
Intemediate 2					
H-bonds[a]	0.41	0.29	0.31	0.50	0.65
C_α contacts[a]	0.57	0.50	0.61	0.79	0.77
SASA hydrophobic[b]	1.23	1.21	1.64	1.06	1.09

[a]Fraction of native hydrogen bonds (H-bonds) and C_α contacts.
[b]SASA of hydrophobic residues relative to the corresponding averaged values of the simulation at 300 K. R1 to R4 denote repeat 1 to repeat 4. The fraction of native interrepeat contacts are not listed because of their high standard deviations.

SASA of hydrophobic residues in the second repeat compared with the TSE [Tables III and IV(b)] indicates that the two N-terminal repeats might have partially repacked against each other. In the I_2 conformations, the first repeat is partly or almost completely unstructured and very flexible (Fig. 10). In particular, the first helix of the N-terminal repeat loses its secondary structure earlier than the second helix in five of six simulations. The second repeat is more unstructured than in the first intermediate. In some conformations of I_2, the fourth helix loses its secondary structure and presents a C_α RMSD from the native state larger than 3 Å. The C-terminal repeats have on average a slightly reduced number of native contacts [Table IV(b)] in I_2. They are almost completely folded in two of six simulations (during 50 and 24 ns, respectively), whereas in the other four runs they lose part but not all of their secondary and tertiary structure [Fig. 11(b)]. The simulation results suggest the presence of a further free energy barrier dividing I_2 from the fully denatured state. The stability of repeats 3 and 4 on a 100-ns timescale at 400 K is consistent with the experimental observation that the two C-terminal repeats were identified as the minimum folding unit of p16.[11] In addition, at least one intermediate state with a reduced Φ-helical content relative to the native state was observed in refolding experiments of p16.[6] The presence of on-pathway partially folded intermediates is consistent with the lack of sequence-distant tertiary contacts in the folded state of ankyrin repeat proteins and their modular topology.

DISCUSSION

Ankyrin repeat proteins differ from globular proteins mainly in two aspects of their native state. First, their modular topology implies that native contacts are present only for residues relatively close along the primary structure. Second, they possess an elongated, essentially toroidal, hydrophobic core that spans from the N-terminal to the C-terminal repeat. Both of these structural properties make ankyrin repeat proteins pertinent targets for experimental[13] and computational studies of protein folding.[32] Here, the structural plasticity of the tumor suppressor p16, whose ankyrin fold consists of four repeats, was investigated by MD simulations at room temperature. On the nanosecond timescale, the loops are more flexible than the helix bundles and the N-terminal half less rigid than the C-terminal half. In particular, the first part of the second repeat, which in the folded state deviates from the canonical ankyrin fold, fluctuates significantly and more than expected from previous NMR analysis. Experimental Φ-value analysis indicates that the second repeat of p16 is unstructured in the TSE. The high flexibility of the N-terminal half of p16 is consistent with the marginal thermodynamic stability and tendency to aggregate[2,5,6] which makes p16 vulnerable to mutations such as those in cancerous tissues.[3,4,6]

The unfolding process of p16 was analyzed by six 100-ns MD simulations at 400 K. It has been shown previously for other proteins that the sequence of unfolding events is essentially the same at physiologic and high temperature and the latter has been used to investigate the folding process.[12,34] However, Dinner and Karplus[35] have observed for a 125-bead protein model on a lattice that the unfolding trajectories at elevated temperature are representative of the folding trajectories below the melting temperature but less diverse. Their lattice simulation analysis indicates that unfolding simulations are most likely to provide useful information for proteins that lack off-pathway intermediates. Thus, increasing the temperature is a very efficient way to accelerate the unfolding process for obtaining more statistics but comparison with experimental data is required to identify eventual artifacts. In the present work, by combining the available experimental Φ values with structural analysis of the MD trajectories, it has been possible to identify snapshots belonging to the TSE, allowing its structural characterization at a level of detail that cannot be achieved by experimental Φ-value analysis alone. The TSE of p16 is rather compact and native-like except for the first part of the second repeat which already in the native state is less structured and fluctuates more than the rest of the protein as mentioned above. The first repeat is nearly intact in the TSE despite its being slightly displaced with respect to the remainder of the structure. Furthermore, significant non-native interactions involving several hydrophobic side-chains have been observed at the TSE. These simulation results supplement the picture obtained by experimental Φ-value analysis, which lacks the description at atomic level of detail.

The presence of partially folded intermediate states along the (un)folding pathways of p16 and repeat proteins in general is compatible with their structural modularity and the near absence of interactions between residues far apart in the sequence.[36,37] Additional evidence on intermediates comes from two unrelated experimental studies on repeat proteins. On the basis of equilibrium denaturation experiments, the (un)folding of the *Drosophila* Notch receptor was suggested to follow multiple pathways with the presence of partially folded conformations consisting of a mixture of structured and unstructured repeats.[33] A designed protein containing three ankyrin repeats showed a two-state behavior during urea-induced denaturation at 5°C, but differential scanning calorimetry experiments

revealed a deviation from two-state model at elevated temperatures.[38] Equilibrium intermediates with reduced α-helical content became populated by stepwise temperature increase from 5° to 65°C.[38] It would be interesting to further validate the probable link between structural modularity and presence of intermediates by refolding experiments of other natural repeat proteins for which thermal and urea denaturation experiments suggested a two-state behavior. For example, no intermediate states of myotrophin could be detected from spectroscopic data,[39,40] but scanning calorimetry experiments were not performed because thermally unfolded myotrophin tends to aggregate at high concentrations.[39] It is important to note that kinetic intermediates are often not sufficiently populated to be detected in unfolding experiments. For this reason, a two-state unfolding behavior of p16 had been assumed in the interpretation of the experimental data.[6,13] However, the presence of partially folded conformations in the MD trajectories indicates that there is more than one free-energy barrier during unfolding. As in previous computational studies by others,[12,15] the assumption has been made here that the experimentally determined Φ values describe the highest of these barriers, i.e., the TSE.

Finally, it is interesting to note that there are large differences in the location of the TSE on the folding reaction coordinate for different repeat proteins. This finding is related to the fact that repeat proteins show a correlation between their thermodynamic stability and the number of repeats.[5,33,37] For example, p18, another ankyrin repeat protein from the same INK family as p16 but with an additional repeat, is thermodynamically more stable (2.98 versus 1.94 kcal/mol at 293 K).[5] The seven-repeat protein *Drosophila* Notch receptor has an even higher stability (6.65 kcal/mol at 288 K) and a very early folding TSE, i.e., close to denatured state, with only one of seven repeats structured.[33] In contrast, p16, which consists of four repeats, has a late folding TSE.[13] The shift of the folding TSE toward the native structure upon stabilization of the denatured state, i.e., reduction of number of repeats in modular proteins, is consistent with the Hammond effect (the movement of a transition state along a reaction coordinate away from a reactant that is stabilized).[41] In other words, the correlation between thermodynamic stability and number of repeats[5] implies that the smaller the latter the closer to the folded state is the TSE.

ACKNOWLEDGMENTS

The authors thank Drs. I. Jelezarov, A.G. Abebe, and S. Cranz for interesting discussions about experimental Φ-value analysis and transition states. The authors are very grateful to Dr. G. Onstad for reading, and English corrections to, the manuscript. The simulations were performed on the Matterhorn Beowulf cluster at the Computing Center of the University of Zurich. The authors thank C. Bolliger, T. Steenbock, and Dr. A. Godknecht for setting up and maintaining the cluster. Preliminary simulations were run on the zBox supercomputer at the Department of Theoretical Physics of the University of Zurich and the authors also thank Dr. J. Stadel and Prof. B. Moore for giving them the opportunity to use the zBox.

REFERENCES

1. Andrade MA, Perez-Iratxeta C, Ponting CP. Protein repeats: structures, functions, and evolution. J Struct Biol 2001;134:117–131.
2. Tevelev A, Byeon IJ, Selby T, et al. Tumor suppressor p16(INK4A): structural characterization of wild-type and mutant proteins by NMR and circular dichroism. Biochemistry 1996;35:9475–9487.
3. Lukas J, Parry D, Aagaard L, et al. Retinoblastoma-protein-dependent cell-cycle inhibition by the tumor-suppressor. Nature 1995;375:503–506.
4. Serrano M. The tumor suppressor protein p16 (INK4a). Exp Cell Res 1997;237:7–13.
5. Yuan C, Li J, Selby TL, Byeon IJ, Tsai MD. Tumor suppressor INK4: comparisons of conformational properties between p16(INK4a) and p18(INK4c). J Mol Biol 1999;294:201–211.
6. Tang KS, Guralnick BJ, Wang WK, Fersht AR, Itzhaki LS. Stability and folding of the tumour suppressor protein p16. J Mol Biol 1999;285:1869–1886.
7. Byeon IJ, Li J, Ericson K, et al. Tumor suppressor p16(INK4A): determination of solution structure and analyses of its interaction with cyclin-dependent kinase 4. Mol Cell 1998;1:421–431.
8. Yuan C, Selby TL, Li J, Byeon IJ, Tsai MD. Tumor suppressor INK4: refinement of p16(INK4A) structure and determination of p15(INK4B) structure by comparative modeling and NMR data. Protein Sci 2000;9:1120–1128.
9. Li J, Byeon IJ, Ericson K, et al. Tumor Suppressor INK4: determination of the solution structure of p18(INK4C) and demonstration of the functional significance of loops in p18(INK4C) and p16(INK4A). Biochemistry 1999;38:2930–2940.
10. Russo AA, Tong L, Lee JO, Jeffrey PD, Pavletich NP. Structural basis for inhibition of the cyclin-dependent kinase Cdk6 by the tumour suppressor p16(INK4a). Nature 1998;395:237–243.
11. Zhang B, Peng ZY. A minimum folding unit in the ankyrin repeat protein p16(INK4), J Mol Biol 2000;299:1121–1132.
12. Daggett V. Molecular dynamics simulations of the protein unfolding/folding reaction. Acc Chem Res 2002;35:422–429.
13. Tang KS, Fersht AR, Itzhaki LS. Sequential unfolding of ankyrin repeats in tumor suppressor. Structure 2003;11:67–73.
14. Fersht AR, Matouschek A, Serrano L. The folding of an enzyme 1. theory of protein engineering analysis of stability and pathway of protein folding. J Mol Biol 1992;224:771–782.
15. Li A, Daggett V. Identification and characterization of the unfolding transition state of chymotrypsin inhibitor 2 by molecular dynamics simulations. J Mol Biol 1996;257:412–429.
16. Jemth P, Gianni S, Day R, et al. Demonstration of a low-energy on-pathway intermediate in a fast-folding protein by kinteics, protein engineering, and simulation. Proc Natl Acad Sci USA 2004;101:6450–6455.
17. Vendruscolo M, Paci E, Dobson CM, Karplus M. Three key residues form a critical contact network in a protein folding transition state. Nature (Lond) 2001;409:641–645.
18. Gsponer J, Caflisch A. Molecular dynamics simulations of protein folding from the transition state. Proc Natl Acad Sci USA 2002;99:6719–6724.
19. 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.
20. Rao F, Settanni G, Guarmera E, Caflisch A. Estimation of protein folding probability from equilibrium simulations. J Chem Phys 2005;122:184901.
21. 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.
22. Hubner IA, Shimada J, Shakhnovich EI. Commitment and nucleation in the protein G transition state. J Mol Biol 2004;336:745–761.
23. Kalé L, Skeel R, Bhandarkar M, et al. NAMD2: greater scalability for parallel molecular dynamics. J Comput Phys 1999;151:283–312.
24. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM—A program for macromolecular energy, minimization, and dynamics calculations. J Comput Chem 1983;4:187–217.

25. Darden T, Darrin Y, Pedersen L. Particle mesh ewald—an n.log(n) method for ewald sums in large systems. J Chem Phys 1993;98: 10089–10092.
26. Serrano L, Matouschek A, Fersht AR. The folding of an enzyme 3. structure of the transition-state for unfolding of barnase analyzed by a protein engineering procedure. J Mol Biol 1992;224:805–818.
27. Sanchez IE, Kiefhaber T. Origin of unusual Phi-values in protein folding: evidence against specific nucleation sites. J Mol Biol 2003;334:1077–1085.
28. Fersht AR, Sato S. Phi-Value analysis and the nature of protein-folding transition states. Proc Natl Acad Sci USA 2004;101:7976–7981.
29. Geierhaas CD, Paci E, Vendruscolo M, Clarke A. Comparison of the transition states for folding of two Ig-like proteins from different superfamilies. J Mol Biol 2004;343:1111–1123.
30. Myers JK, Pace CN, Scholtz JM. Denaturant m-values and heat-capacity changes—relation to changes in accessible surface-areas of protein unfolding. Protein Sci 1995;4:2138–2148.
31. Kobe B, Deisenhofer J. A structural basis of the interactions between leucine-rich repeats and protein ligands. Nature 1995;374: 183–186.
32. Ferreiro DU, Cho SS, Komives EA, Wolynes PG. The energy landscape of modular repeat proteins: topology determines folding mechanism in the ankyrin family. J Mol Biol 2005;354:679–692.
33. Mello CC, Barrick D. An experimentally determined protein folding energy landscape. Proc Natl Acad Sci USA 2004;101:14102–14107.
34. Cavalli A, Ferrara P, Caflisch A. Weak temperature dependence of the free energy surface and folding pathways of structured peptides. Proteins 2002;47:305–314.
35. Dinner AR, Karplus M. Is protein unfolding the reverse of protein folding? A lattice simulation analysis. J Mol Biol 1999;292:403–419.
36. Tripp KW, Barrick D. The tolerance of a modular protein to duplication and deletion of internal repeats. J Mol Biol 2004;344: 169–178.
37. Kajander T, Cortajarena AL, Main ER, Mochrie SGJ, Regan L. A new folding pradigm for repeat proteins. J Am Chem Soc 2005;127: 10188–10190.
38. Devi SV, Binz HK, Stumpp MT, Plückthun A, Bosshard HR, Jelesarov I. Folding of a designed simple ankyrin repeat protein. Protein Sci 2004;13:2864–2870.
39. Mosavi LK, Williams S, Peng Z. Equilibrium folding and stability of myotrophin: a model ankyrin repeat protein. J Mol Biol 2002;320:165–170.
40. Mosavi LK, Cammett TJ, Desrosiers DC, Peng Z. The ankyrin repeat as molecular architecture for protein recognition. Protein Sci 2004;13:1435–1448.
41. Hammond GS. A correlation of reaction rates. J Am Chem Soc 1955;77:334–338.
42. Kabsch W, Sander C. MOLMOL: a program for display and analysis of macromolecular structures. Biopolymers 1983;22:2577–2637.
43. Koradi R, Billeter M, Wuthrich K. Dictionary of protein secondary structure—pattern-recognition of hydrogen-bonded and geometrical features. J Mol Graph 1996;14:29–32, 51–55.