# Unfolding dynamics of the protein ubiquitin: Insight from simulation

Shubhra Ghosh Dastidar and Chaitali Mukhopadhyay*
*Department of Chemistry, University of Calcutta, 92 A. P. C. Road, Kolkata—700 009, India*
(Received 26 May 2005; published 29 November 2005)

The temperature-induced unfolding pathway of ubiquitin has been investigated by molecular dynamics simulation at four different temperatures. It has been observed that the sequences of the unfolding events are same at all the temperatures. However, the time scale of the dynamics at different temperatures are different. The transition states at various temperatures also possess similar secondary structural elements. The intermediate conformations visited by the protein at different temperatures can help determination of the transition states. The well known "*A* state" of ubiquitin, hitherto found to be stable only in methanol water mixture, have been observed to be a common transient intermediate conformation in the unfolding path of the protein in water. Our observation about the similarities of the unfolding process at different temperatures strongly recommend for a defined pathway for the unfolding process.

DOI: 10.1103/PhysRevE.72.051928 PACS number(s): 87.14.Ee, 87.15.He

## I. INTRODUCTION

Understanding the mechanism of the folding/unfolding processes of the naturally occurring proteins has remained one of the major challenges to the scientists due to the remarkable fact that proteins fold reliably and quickly to their native state despite the astronomical number of possible conformations and this puzzling observation is well known as the "Levinthal paradox" [1]. Much of the insight about the folding-unfolding process has been obtained from various experimental data and has been well complemented by theoretical results [2–24]. Of the available theoretical methods, molecular-dynamics (MD) simulation is particularly useful to provide atomic level details of a large range of processes associated with protein motion on a picosecond to microsecond time scale [5]. Several recent reports on simulation of protein unfolding have indicated that such simulations improve the possibility of searching the relevant intermediate conformations that are otherwise difficult to characterize experimentally due to their poor stability [6]. The folding simulation requires very long simulation run time and has very low success rate, whereas unfolding simulation starts from the native structure of a protein and the entire reaction coordinate starting from the native to the denatured state could be explored in a finite MD simulation runtime and significant amount of information about the intermediate/transition states as well as the overall process could be obtained [6–9]. The structure of the transition state (TS) might provide a wealth of information. It has been proposed recently that formation of a few correct contacts between a set of key residues are very crucial for the folding [25] of the proteins and these contacts are retained in the transition state ensemble [10]. To understand the mechanism of folding it is thus necessary to identify these contacts and the processes of their formation/disruption leading to the folding/unfolding of the protein.

Ubiquitin is a small globular protein known to fold in a two-state manner [11–14,26–29]. The structure of the protein is highly stable at neutral pH, the melting temperature ($T_m$) is ~373 K and it decreases with the decrease in the pH [11]. A recent report suggested that the elucidation of the structure of the transition state (TS) in the folding pathway depends on the method of calculation [29] (mutational $\phi$ or $\psi$ value) and two different structures of the TS for ubiquitin can be proposed. The first was proposed from $\psi$-value analysis, which contained four $\beta$ strands, one $\alpha$ helix, whereas another proposed $\phi$-value analysis contained only a $\beta$ hairpin and the $\alpha$ helix; let them be referred to as TS1 and TS2, respectively. The structure of TS1 is very close to the reported molten globule structure of ubiquitin, which is stable in 60% methanol–40% water, but is yet to be characterized as an intermediate in pure water [13,14]. On the other hand, a structure almost similar to TS2 has been reported from high resolution NMR [28] as thermally stable even at pH 2. But there is no experimental evidence for the presence of multiple transition states for ubiquitin. A recent report from time resolved NMR study, has suggested that the folding-unfolding behavior of ubiquitin does not fit to a two state equilibrium [12]. This observation indicates the presence of a transient intermediate state, which has not been very clearly identified. We have investigated the unfolding steps of ubiquitin at different temperatures to compare the similarities and dissimilarities as well as to identify the intermediate conformational states that the protein might visit during the unfolding process.

## II. METHOD

The native structure of the ubiquitin was obtained from the protein data bank (PDB code 1UBQ). The protein was then placed in a cubic water box of water density 0.98 gm/cc. The TIP3P [30] water model was used. The water molecules whose oxygen atoms did overlap (within 2.6 Å) with the heavy atoms of the protein were deleted and 2882 number of water molecules survived. Then the system was the well minimized. CHARMM22 force field and parameters [31] were used for all the calculations. The resulting system was equilibrated at four different temperatures inde-

*Corresponding author. Email address: chaitalicu@yahoo.com

 

TABLE I. Summery of the trajectories prepared for entire work indicating the temperatures, trajectory length, fate of the protein structures, and the densities of water.

| Trajectory No. | Temperature | Trajectory length (ns) | Fate of the protein structure | Density of water (gm/cc) |
|---|---|---|---|---|
| I | 300 K | 6.0 | Native | 0.98 |
| II | 385 K | 18.0 | Unfolded | 0.86 |
| III | 450 K | 8.8 | Unfolded | 0.81 |
| IV | 520 K | 1.0 | Unfolded | 0.73 |

pendently keeping the atmospheric pressure as a constant parameter. The system equilibrated at 300 K had the dimension 48 Å $\times$ 48 Å $\times$ 48 Å, but at the higher temperatures the water box expanded leading to the following dimensions: 50 Å $\times$ 50 Å $\times$ 50 Å at 385 K, 51 Å $\times$ 51 Å $\times$ 51 Å at 450 K, and 53 Å $\times$ 53 Å $\times$ 53 Å at 520 K. The resulting densities of water were 0.98 gm/cc at 300 K, 0.86 gm/cc at 385 K, 0.81 gm/cc at 450 K, and 0.73 gm/cc at 520 K. Similar ranges of the value for the densities of water at high temperatures have also been reported previously [7,32,33]. We choose the neutral pH because ubiquitin has a very significant stability at this pH.

NPT simulation run was performed at different temperatures mentioned above. Each trajectory was truncated after a sufficient progress of unfolding process except for the control run at 300 K, which did not show any unfolding and it was truncated after 6 ns run. SHAKE was applied to freeze the vibration of the bonds which allowed the use of 2 fs time step for the integration. The particle mesh Ewald (PME) technique was used to handle the long-range interactions [34]. The nonbonded lists were updated after each 25 steps and the coordinates were saved after each 1 ps.

## III. RESULTS

Table I summarizes the MD simulation run that has been carried out. At 300 K the structure of the protein remained very close to that of the crystal structure, shown in Fig. 1(a). At 300 K, ubiquitin retains the native secondary structures including five $\beta$ strands between residues 1–8($\beta$1), 11–17($\beta$2), 40–45($\beta$3), 48–50($\beta$4), and 64–72($\beta$5), that construct the $\beta$ sheets with the following interactions: $\beta1/\beta2$, $\beta1/\beta5$, $\beta5/\beta3$, $\beta3/\beta4$, where the slashes indicate an interface. $\beta1/\beta2$ and $\beta3/\beta4$ are two $\beta$ hairpins. In addition, there are two helical structures ranging from the residue numbers 24–34 ($\alpha$ helix) and 56–59 ($3_{10}$ helix). This trajectory (trajectory I) was truncated after 6 ns simulation run since no indication of the unfolding was observed. The average structure of this trajectory represented the native state of ubiquitin. The structural and dynamical properties of the protein obtained from this trajectory have been used as the reference data. In a pervious work [35] we found the exactly same structure of ubiquitin when it was simulated in water at 300 K.

### A. Dynamics of the backbone

The time dependant variation of the root mean square deviation (RMSD) (Fig. 2) of the backbone atoms ($C_\alpha$ and N atoms) of the protein from the starting structure gives an estimation of the rate of the unfolding at different temperatures. The control run at 300 K had an average RMSD of $\sim$1.5 Å for the backbone atoms. The melting temperature ($T_m$) of ubiquitin is $\sim$373 K at neutral pH [11]. In the unfolding trajectory run at 385 K, which is just above the $T_m$, the protein started to unfold after 2 ns, as obvious from the rise of RMSD from $\sim$2 to $\sim$5 Å between the time points 2 and 3 ns. During this rise in RMSD, the protein loses the most "soft" tertiary contacts, which has been discussed in the later parts of this article. At this temperature the protein required $\sim$18 ns time for complete denaturation (RMSD $\sim$12 Å). At 450 K, the triggering of the unfolding showing a sharp increase of RMSD was observed between $\sim$1 to $\sim$2 ns and the same at 520 K was observed between $\sim$200 to $\sim$300 ps. As the protein reached the RMSD of the range $\sim$5–7 Å the rate of change of RMSD is diminished and such feature is common at all the temperatures. The lower the temperature the higher is the duration for which the system remains trapped around a particular range of RMSD. At 385 K, between 3 to 13 ns the RMSD values fluctuates roughly within 6$\pm$1.5 Å, at 450 K this happened between the time points 2.5 to 5.0 ns, whereas at 520 K it

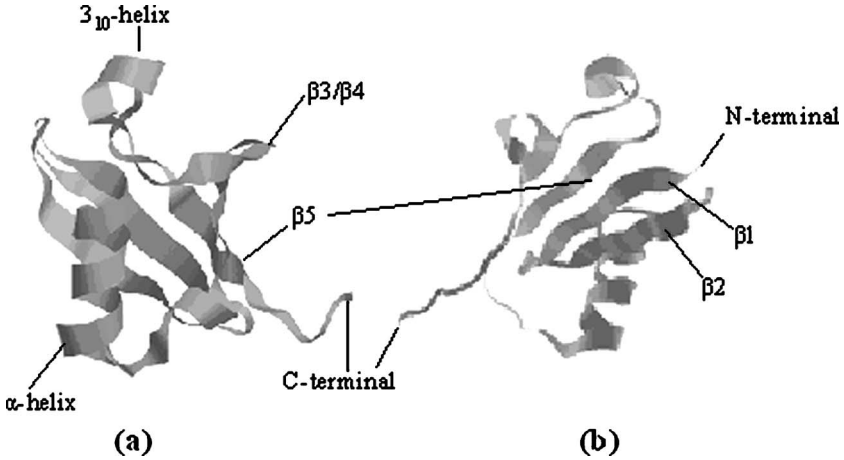


FIG. 1. (a) The native structure of ubiquitin. (b) Structure of one representative of the transition state ensemble.

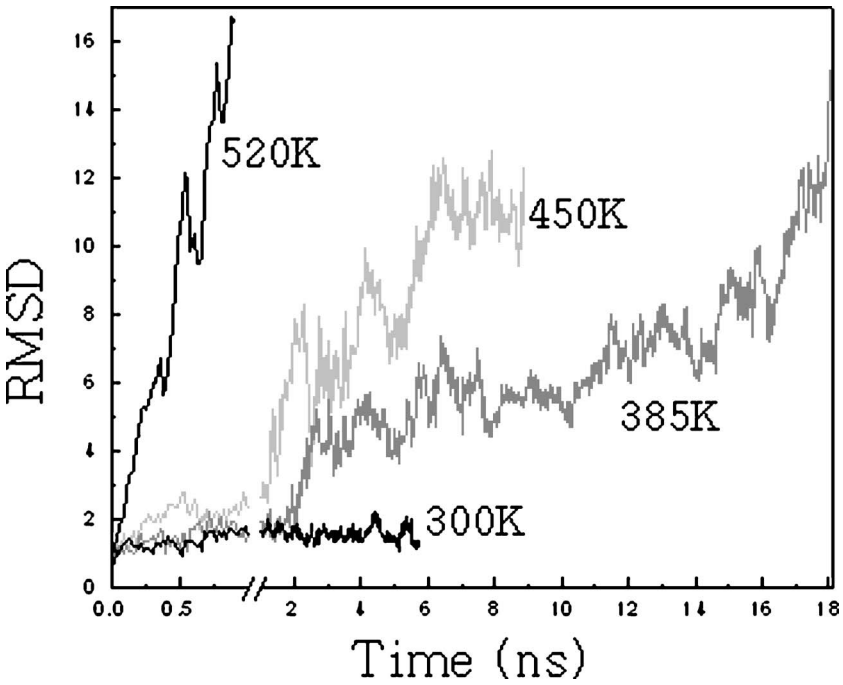


FIG. 2. RMSD vs time plot of the backbone of the protein at different temperatures. From left to right: 520 K (black), 450 K (light gray), at 385 K (gray), 300 K (black). The temperatures have also been shown in the figure. The increment of the scale for the horizontal axis has been set differently so that the rapid change at 520 K could be viewed clearly.

lasted during 300 to 400 ps. During this time intervals either similar conformations or different conformations of similar RMSD vales have been generated. Beyond 8 Å, there is again a sharp increase of RMSD value up to 10 to 12 Å. At this point $\beta 1/\beta 5$ interaction was lost which resulted in the complete destruction of the tertiary structure of the protein.

There was indication from the previous experimental report [14] that the secondary and tertiary structure constructed by the residues 1–33 is comparatively rigid compared to the other half of the protein (i.e., residue Nos. 34–76) and the residues 1–33 almost retains their native topological properties in the "*A* state" of Ubiquitin. To investigate such a possibility, the RMSD of the N terminal 1–33 residues has been separated from the residues 34–76 and both have been plotted in Fig. 3. Data has been shown only for two temperatures 385 and 520 K. At 450 K similar trends of the changes have been observed (data not shown). The Fig. 3 shows that previous experimental report is in full agreement with the simulated data. At both the temperatures a rapid denaturation takes place at the C-terminal residues (34–76), whereas the N-terminal residues denature at a much slower rate. The variation of RMSD of the core defined by the hydrophobic residues M1, I3, V5, T13, L15, V17, and V26 of the protein also indicates the retention of their positions at 385 K for about 10 ns after which there is a gradual increase in RMSD. A recent report [15] of the study on the effect of mutation on the stability and folding/unfolding rate of the ubiquitin indicates that the variation of the length of the side chains of

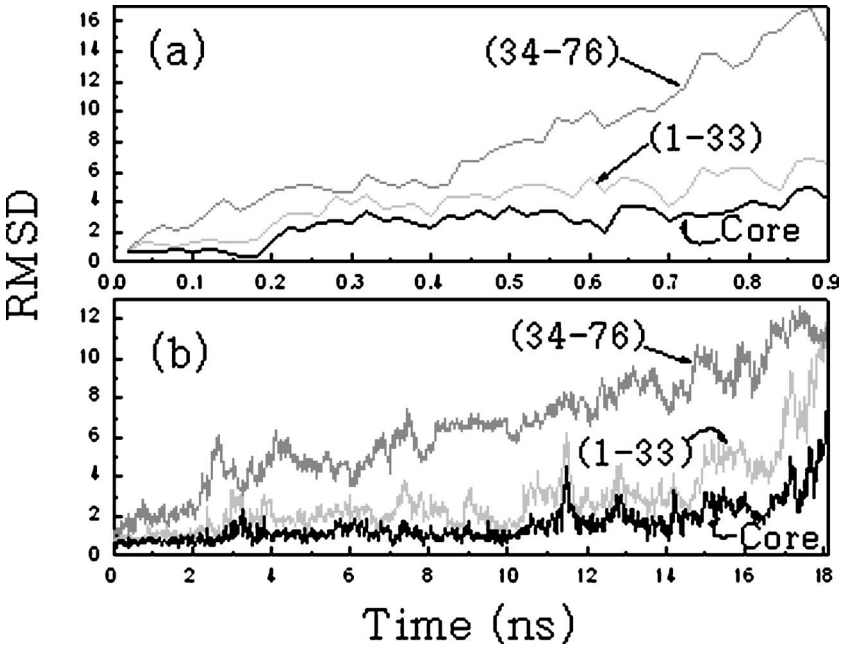


FIG. 3. RMSD vs time plot of the different portion of the backbone of the protein: (a) at 520 K and (b) at 385 K. The number within the parentheses indicate the data for a range of residues. The coloring scheme is as follows: residue No. 1–33 (light gray), 34–76 (gray), core residues (black). The definition of the core residues has been mentioned in the text.

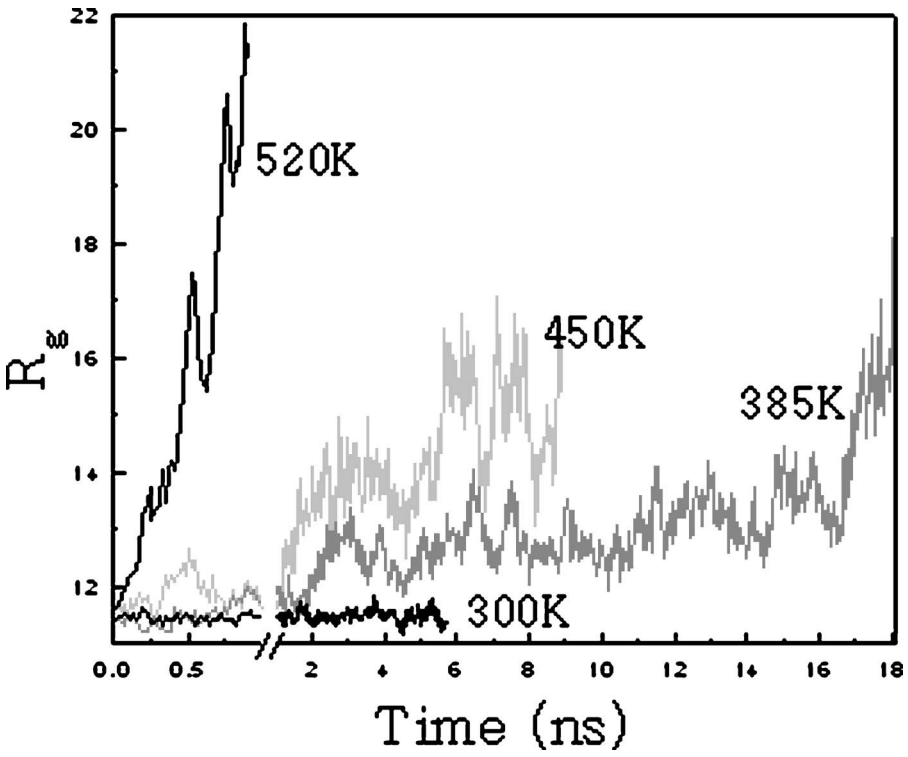


FIG. 4. Plot of radius of gyration ($R_g$) of the protein vs time at different temperatures. The coloring scheme and the increment on the scale in the horizontal axis is similar to that in Fig. 2.

theses core residues greatly affect the stability of the mutated structure. The least variation of the position of these residues shown in Fig. 3 also implies that the topology formed by these residues are very important for resisting the destruction of the structure.

### B. Increase of radius and surface area due to conformational changes

The change of the radius of the globular protein ubiquitin could be a good estimation of the extent of denaturation of the protein structure. The plot of radius of gyration ($R_g$) of the protein vs time (Fig. 4) shows similar trends at different temperatures. The native protein at 300 K has a lowest value of $R_g$ around ~11 Å and that value exceeds 16 Å for completely denatured protein. At 520 K the $R_g$ crossed the value of 16 Å (trajectory IV) within 500 ps and at 450 K the same incident happened around 6 ns. At 385 K there was a sudden jump of $R_g$ value from 14 to 16 Å after 17 ns. This sharp change was common to all the unfolding trajectories and as discussed later, this was due to the complete loss of $\beta 1/\beta 5$ interaction. The ensemble of the structures having $R_g$ within 15–16 Å has only the N-terminal $\beta$ hairpin and the $\alpha$ helix as two isolated secondary structural units and completely devoid of any 3D contacts between these two secondary structural units. The further drastic increase of $R_g$ at 520 K was due to loss of the $\beta$-hairpin structure of the N terminus, a residual part of which was still retained at 450 K or lower temperatures at the end of those trajectories. The overall variation of the $R_g$ at all the temperatures is same in nature with that of backbone RMSD.

The opening of the hydrophobic core of the protein and the reaching of the water molecules into the core can accelerate the denaturation process. The increase of the solvent accessible surface area (SASA) with time has been shown in Fig. 5. SASA has been calculated with a spherical probe of radius 1.6 Å. At 300 K the SASA remains unchanged and its value at this temperature indicates that the accessible surface area of native protein is about 5000 Å$^2$. Both at 385 and 450 K the fast initial increase of SASA is followed by a comparatively slow change. The initial destruction of the "soft" tertiary contacts of the protein allows the water to come into the core of the molecule, which is reflected by the initial fast increase of the SASA, and its further increase results from the slow destruction of the 3D scaffold maintained by comparatively rigid part of the protein. The large value of SASA, which has been observed at 520 K, was due to the near-complete destruction of the tertiary scaffold. Similar to $R_g$, the SASA also has the minimum value at the native conformation.

### C. Tertiary interactions

A quantitative estimation of the compactness of a particular conformation of the protein could be obtained from the calculation of the total number of tertiary contacts ($N_{\text{cont}}$) present in the structure. A tertiary contact is counted if the centers of mass of the side chains of two different residues come within a distance of 5.4 Å or any two atoms of different side chains come within a distance of 4.6 Å [7]. Figure 6 shows the variation of the total number of such contacts as a function of time at different temperatures. Number of such native contacts for the native protein is ~120 but for a completely denatured protein this number falls to ~60 (at 520 K), which was observed at the end of the trajectory IV. Since at the end of the trajectory III at 450 K the final structure contained the $\alpha$ helix and a very short portion of the $N$-terminal $\beta$ hairpin, at this temperature the minimum value

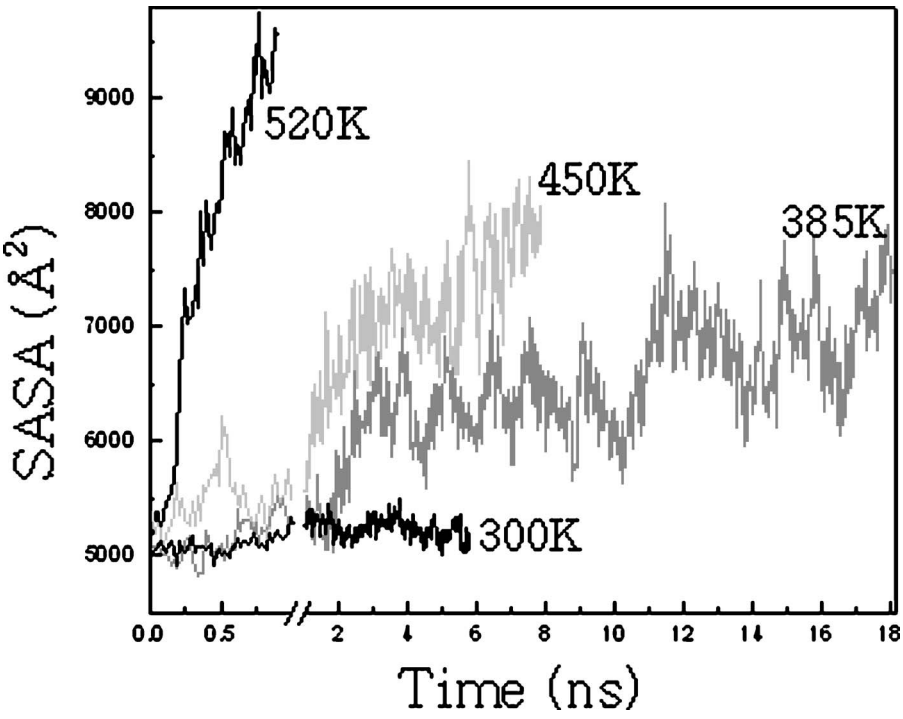


FIG. 5. Change of solvent accessible surface area (SASA) of the protein with time at different temperatures. The coloring scheme and the increment on the scale in the horizontal axis is same with the Fig. 2.

of $N_{cont}$ was ~80, and this was also true for trajectory II at 385 K. The side chains of the surface residues are more highly mobile than the backbone (due to the slow denaturation process) of the protein and so there are always the high fluctuations of the values of $N_{cont}$.

The plots of $C_\alpha$-$C_\alpha$ contacts at various time points at different temperatures are shown in Figs. 7(a)–7(h), which gives the information about the protein skeleton at different time points. If the $C_\alpha$ atoms of $i$th and $j$th atom are within 5.5 Å then it has been shown with a black point at the coordinate $(i,j)$ in the map [Figs. 7(a)–7(h)] and only the half diagonal part of this symmetric plot has been shown. The structures have been averaged within a window of ±25 ps around the time point of the structure mentioned. As was observed from the residue-wise RMSD plots (Fig. 3), the process of the unfolding at the early stages was not distributed evenly over the entire protein. Since the unfolding rate is slowest at 385 K, it is easier to get information about the details of the early stages of unfolding from this trajectory. The initial loss of the tertiary contacts was first observed

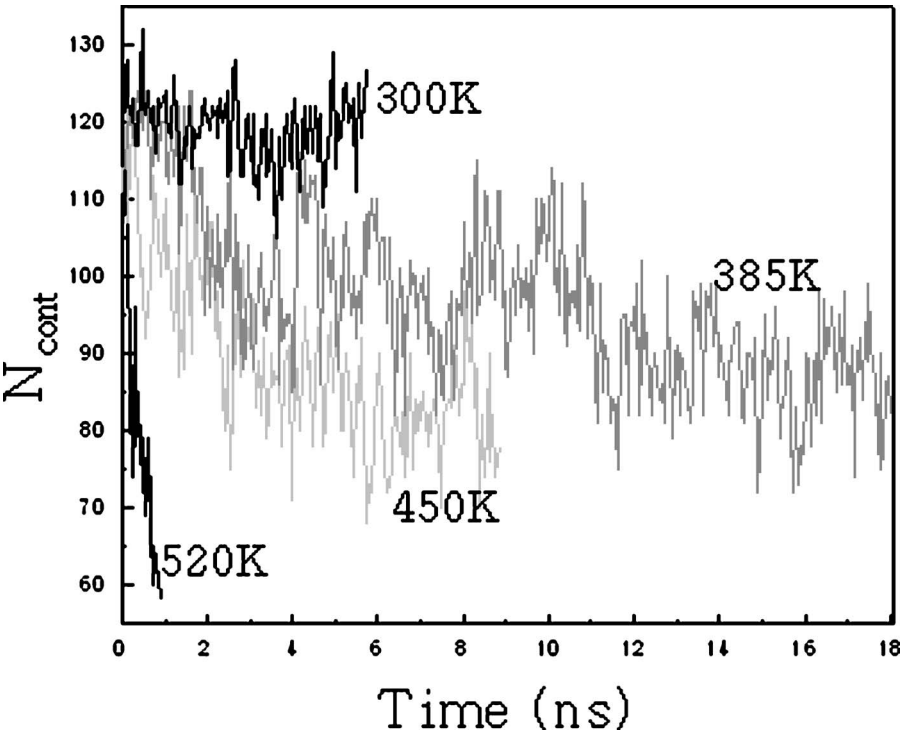


FIG. 6. Time-dependent variation of the tertiary contact ($N_{cont}$) of the protein at different temperatures. The coloring scheme is similar to the Fig. 2.

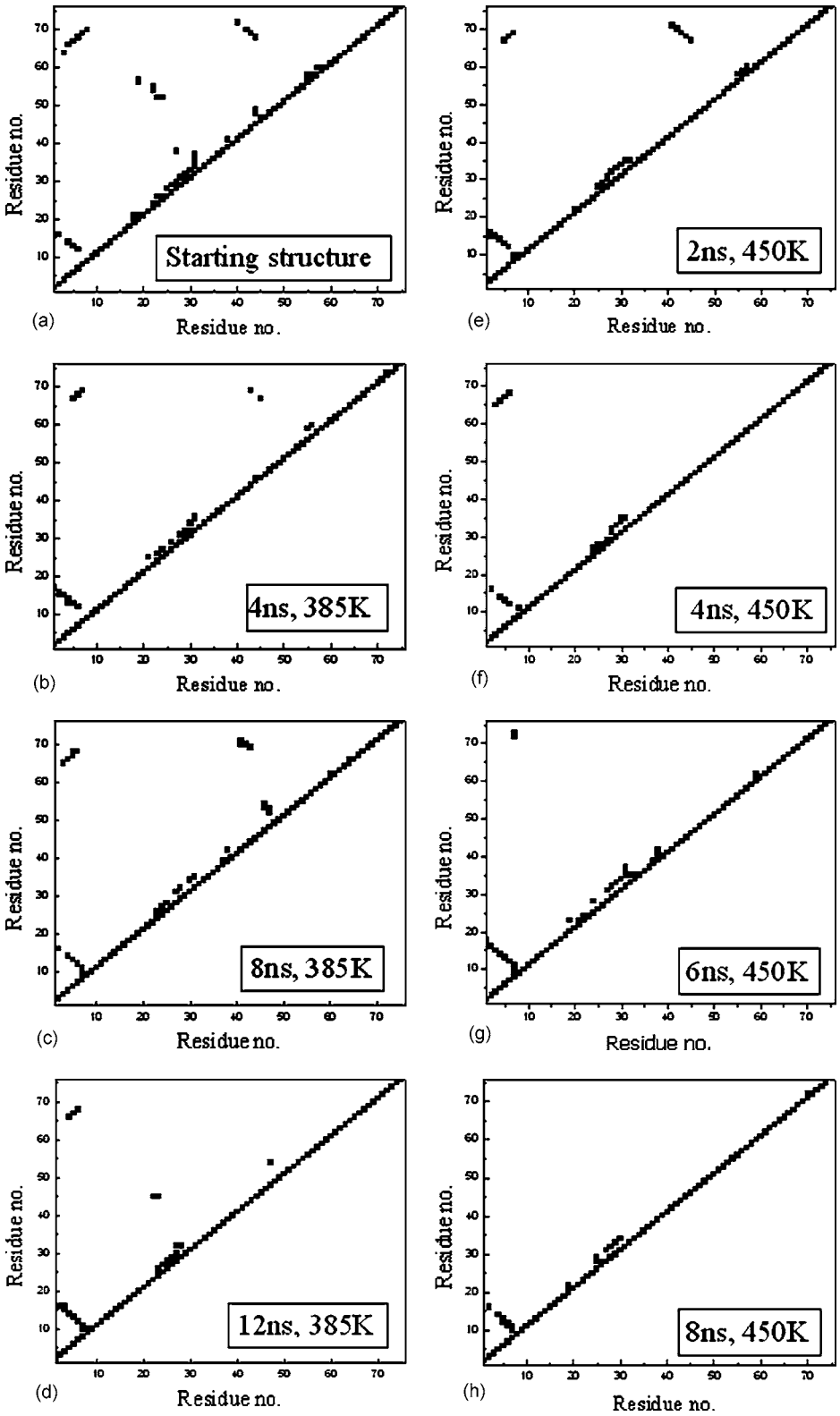


FIG. 7. $C_\alpha$-$C_\alpha$ contact map of the protein at different temperatures and at different time points. (a) Starting structure, (b) after 4 ns, at 385 K, (c) after 8 ns, at 385 K, (d) after 12 ns, at 385 K, (e) after 2 ns, at 450 K, (f) after 4 ns, at 450 K, (g) after 6 ns, at 450 K, (h) after 8 ns, 450 K.

between the two sets of residues, 19–23 and 52–56. These most "soft" contacts are between the *N* termini of the $\alpha$ helix and the $3_{10}$ helix (residues 56–59). The loss of contact between these residues makes it easier for the opening of the core of the protein in the later steps. Next, the contacts between the interfaces $\beta 3/\beta 4$ and $\beta 3/\beta 5$ are weakened [Figs. 7(b) and 7(c), sequentially], followed by complete loss of the $3_{10}$ helix [Fig. 7(d)]. The contacts between $\beta 5$ and $\beta 1/\beta 2$ are retained even after 12 ns at 385 K. It is interesting to note that simulations at higher temperatures agree with this sequence of events. The $C_\alpha$-$C_\alpha$ plots at 450 K [Figs. 7(e)–7(h)] show that the long lasting secondary structural elements of the protein are the $\alpha$ helix and the N-terminal $\beta$ hairpin. Exactly the same trend has been observed at 520 K

(data not shown) and the existence of the $\beta$ hairpin at 520 K even after 600 ps indicates a very high but unusual stability, which has been also observed from the experiment [28]. An important observation is that at all the temperatures there is no evidence for the formation of non-native tertiary interaction during the course of unfolding. If we take points in Fig. 7(a) (starting structure) as the reference we only see the disappearance of points in the other [Figs. 7(b)–7(h)] plots rather than appearance of any new points. The $C_\alpha$-$C_\alpha$ contact matrix shows that the sequence of the events is more or less same at all the temperatures.

### D. Conformational analysis and characterization of the transition state

Transition state (TS) is the highest energy barrier that is present in the pathway of a kinetic process. The transition state is kinetically and thermodynamically unstable and it is expected that the structure of the protein should change rapidly once it passes the major transition state, and this fact was suggested earlier [36]. The identification of the conformation from which the rapid structural changes began to occur could be a guide to identify the transition state. In the time vs RMSD plot of the protein backbone (Fig. 2) two structures having similar RMSD values may have different conformations. So from this plot it is difficult to identify the points where the rapid change of conformations has initiated. But this could be done by analyzing the RMSD between the all pair of conformation generated in a simulated trajectory. These RMSD values could be shown (or plotted) in a two-dimensional (2D) or three-dimensional (3D) RMSD space that can efficiently represent the RMS difference between any pair of conformations and clearly indicate the rapid conformational changes. Several reports are there for successful identification of the TS from this analysis [7,36–38]. In this method N number of conformations are chosen from a MD simulation trajectory. The $C_\alpha$-RMS deviation between each pairs of conformations are calculated and stored in a matrix, e.g., if the matrix is $D$, then $D_{ij}$ element is the RMS deviation between $i$th and $j$th structures. Now a set of $N$ number of points is to be taken in a two-dimensional space and the distance between $i$th and $j$th points is stored in "$d_{ij}$." For example, if the Cartesian coordinates of $i$th and $j$th points are $(x_i, y_i)$ and $(x_j, y_j)$ then,

$$d_{ij} = \{(x_i - x_j)^2 + (y_i - y_j)^2\}^{1/2}. \tag{1}$$

The $N$ number of points in the 2D space should be chosen in such a way so that the distance between the points "$i$" and "$j$," i.e., "$d_{ij}$" is as close as possible to "$D_{ij}$" (the actual RMSD between the $i$th and $j$th conformation). In other words the RMSD matrix $(D_{ij})$ will be fitted with a set of points in space. The absolute values of the coordinates of the points are not important and so the set points could be chosen at any location of the 2D space, but the relative position of the points are important. The total residual $\xi$ is the following equation is minimized:

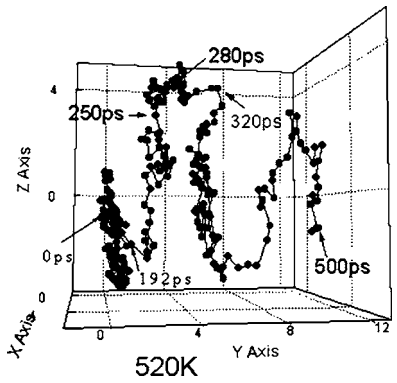


FIG. 8. The three-dimensional plot of the protein conformations in the RMSD space. The distance between two points implies the RMS difference between two conformations. The line connects the two conformations evolved sequentially in time. The resolution is of 2 ps time interval between the conformations.

$$\xi = \sum_{i=1}^{N} \sum_{j>i}^{N} (D_{ij} - d_{ij})^2. \tag{2}$$

The set points in the 2D space, which represents structurally close conformations, will form a cluster in the 2D plot. The two conformations that are structurally dissimilar will remain as separated by a distance, which is very close to the RMS difference of these two structures.

The idea could be extended from 2D to 3D (three dimensions) and addition of an extra dimension will lead to more accurate result. In that case, a set of $N$ number of points is to be taken in a three-dimensional space and the distance between $i$th and $j$th points is stored in $d_{ij}$. For example if the Cartesian coordinates of $i$th and $j$th points are $(x_i, y_i, z_i)$ and $(x_j, y_j, z_j)$ then

$$d_{ij} = \{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2\}^{1/2}. \tag{3}$$

Equation (2), in 3D case can also be used for the best fitting of the data. The $N$ points are chosen in such a way that it can lead to a minimum value of the $\xi$. The minimization of $\xi$ in Eq. (2) will ensure that the $d_{ij}$ is close to $D_{ij}$ as much as possible. We have performed the conformational analysis using a 3D plot shown in Fig. 8 and the coordinates of the $N$ number of points has been shown, where the distance between the points $i$ and $j$, i.e., $d_{ij}$ is as close as possible to $D_{ij}$ (the actual RMSD between the $i$th and $j$th conformation). Each point in the plot represents one individual conformation and a line has connected the conformations, which are evolved sequentially in time.

Figure 8 shows the 3D-RMSD plot of the unfolding trajectory at 520 K. The distance between the points represents the RMS deviation between the conformations. Points are sequentially connected in time and the time difference between two connected conformations is 2 ps. In this type of representation the set of conformations, which are similar or very close to each other will form a cluster. In the three-dimensional representation of the RMSD space at 520 K there are two tightly packed clusters, found within 190 ps. Starting from the native, the cluster expands to RMSD

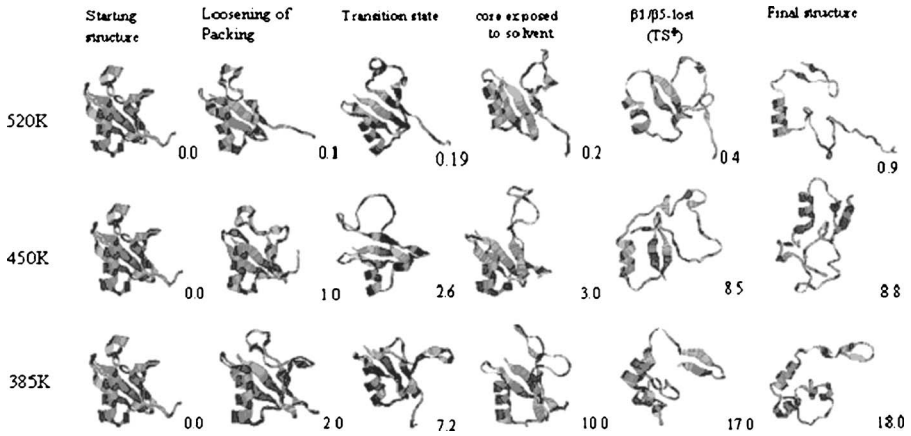


FIG. 9. Snapshots of the protein structure evolved at different time points at different temperatures. The structures having similar secondary and tertiary interactions but generated at different temperatures have been put in a same column. The first column shows the starting structure, second column shows the initial disruption leading to the transition state (third column). The fourth and fifth column are the steps in the way of complete denaturation. The last column show the final structure generated at the time of truncation of the trajectory. The numbers in the right bottom corner of each structure indicate the time point in nanosecond time scale.

4–5 Å. The points in a cluster do not imply that the conformations are identical but the closeness of the points indicates that no rapid conformational changes occurred before 190 ps. The conformation generated at 192 ps initiated the rapid conformational changes and we intend to believe that at this point protein crosses the highest energy barrier and is a likely candidate for the TS ensemble. The structure of the identified TS, looks very close to the proposed TS from the experimental data ($\psi$-value analysis). It contains $\alpha$ helix and four $\beta$ sheets, including the residual part of the $\beta 3/\beta 4$ hairpin. This structure of the TS has been shown in Fig. 1(b). Again after crossing 250 ps the clustering of conformation occurs which continued up to ~312 ps and the conformations in this cluster are very similar to the reported molten globule structure (*A* form) of the protein. This cluster indicates that there could be an existence of a transient intermediate state whose structure is close to the molten globule state of the protein, which is not very stable in the water. The point after the 500 ps have not been considered for this plot since the protein almost denatured within this time period and a better convergence of the fitting of data is possible with lesser number of points. Following the same procedure we have found the transition state at 450 K to be present around 2.6 ns and at 385 K it was around 7.2 ns. The representative conformations of the TS ensembles at different temperatures have been shown in Fig. 9. The residual secondary structures in the TS at all temperatures are similar. It is a disrupted form of the native state and no new tertiary contact (which is not present in the native state) is present. It was suggested in the earlier reports [39] that the major transition state of protein folding/unfolding would be a distorted form of the native state, which meets with our observation. The findings of recent time [10] on the TS ensemble also in full agreement with our results.

## IV. DISCUSSION

The overall unfolding process can be characterized using different parameters. The dynamics of the backbone holds the key information about the unfolding process, presented in Fig. 2. This figure indicates that at 385 K at least 18 ns simulation time is necessary for denaturation, which gets shorter and becomes ~8–9 ns when the temperature is raised to 450 and at 520 K the denaturation is complete in less than 1 ns. But only the backbone dynamics is not enough to understand the process. The collapse of the native structure could increase or decrease its volume. From the change of $R_g$ of the protein (Fig. 4) it is clear that a deviation from the native structure increased the size. Similar to the merging of isolated smaller oil droplets into a larger droplet to decrease the oil-water interfacial interaction, the folded protein minimizes its water accessible surface area in water environment. That is why we see from the Fig. 5 that upon denaturation of the protein SASA is always higher than the native state. The native state is also stabilized by the interactions between the side-chains of the amino acid residues and the disappearance of such interaction as a function of unfolding time has been shown in Fig. 6.

Figure 9 summarizes the unfolding processes at different temperatures. The structures having similar kind of tertiary interaction have been placed in the same column. We have mentioned previously that the loss of tertiary interaction was first observed between the residues 19–23 and 52–56. The comparison of the points within each pairs of Figs. 7(a) and 7(b) and 7(a)–7(e) show the similarities of the initial destruction at the 385 and 450 K. The residues of these two ranges are P19, S20, D21, T22, T23 and D52, G53, R54, T55, L56. The early stages of denaturation included the loss of H-bond between E64→Q2, T22→D52, T23→R54 and also between E24→D52. It has been found from the high-resolution NMR experiments that the H bond between E64→Q2 is most thermolabile among all the intraprotein H bonds [28] and destruction of this H bond makes the $\beta 5$ strand much labile which in turn affects the stability of the $\beta 3/\beta 5$ junction. Our observation matches with the experimental observations. At 520 K also the similar changes has been observed (data not shown). The structures generated after such losses has been put in the second column of the Fig. 9.

The partial destruction of the $\beta3/\beta5$ and $\beta3/\beta4$ sheets and complete destruction of the $3_{10}$ helix continued simultaneously and the backbone of the protein containing these residues was pulled away from the N-terminal part leading to an increased solvent accessibility. Such events lead the protein towards the transition state. We have performed the conformational analysis using a 3D plot in the RMSD space. The TS has almost unaffected $\alpha$ helix, and $\beta1/\beta2$, along with the partially destroyed $\beta1/\beta5$, $\beta3/\beta5$ interaction. The presence of a residual part of the $\beta3/\beta4$ hairpin also observed at 520 K and at 385 K but not observed at 450 K. The structure of the TS matches very well with the structure obtained from the $\psi$-value analysis experiment [29].

The destruction of the $\beta3/\beta5$ interaction leads to the complete exposure of the core to the solvent and for theses conformations the SASA increases to more than 7000 $Å^2$. At this time the denatured protein has $\alpha$ helix, $\beta1/\beta2$, and a residual part of the $\beta1/\beta5$ interaction. This conformation matches with the reported molten globule state (MGS) of the protein stable only in 60% methanol-water mixture [14,40]. Since MGS of ubiquitin is not stable in water we also do not see any significant amount of stability. From the plot shown in Fig. 8 we got an indication that MGS could be present in the unfolding process in water as a transient intermediate state. The destruction of the $\beta1/\beta5$ interaction in the MGS generates the conformations, which have two autonomous units, the major portion of the $\alpha$ helix and almost intact $\beta1/\beta2$ hairpin. But between these two secondary structural elements no tertiary interaction does exist. These two secondary structural elements are the most rigid portions of the ubiquitin structure. The conformations containing only these two autonomous units have been identified as the TS from the $\phi$-value analysis [29]. The $TS^{\phi}$ structures obtained from different trajectories have been placed in the fifth column of the Fig. 9.

At 520 K there is complete loss of $\beta1/\beta2$ hairpin and the final structure after 900 ps simulation run contained only the helix. At comparatively lower temperatures, i.e., at 450 and at 385 K, the residual part of the $\beta$ hairpin was observed for a long time and the trajectories were truncated before their complete destruction. The major difference in the sequence of the unfolding events at different temperatures is that at very high temperatures (450 and 520 K) the loss of the $3_{10}$ helix was faster than that of the $\beta3/\beta4$ and $\beta4/\beta5$ sheets. At 385 K a continuous destruction and reformation of the $\beta$ sheet at the $\beta1/\beta5$ interface was observed and finally that was completely lost around 17 ns and a sudden increase of RMSD (Fig. 2) was observed.

## V. CONCLUSION

In this work we have tried to find out the answer of some questions: Is the unfolding process the same at all the temperatures? What is the unfolding transition state for ubiquitin? Is this consistent at all temperatures? What are the structurally most rigid parts of this protein? Is there any portion on the protein that is crucial to maintain the 3D scaffold of the protein? Do the experiments support the findings?

In the last section we discussed that the steps of the unfolding process are more or less same at all the temperatures including the transition states. The $\alpha$ helix is the most stable part of the protein, but we also have seen an unusual high thermal stability of N-terminal $\beta$ hairpin and there is experimental evidence for such stabilities [28]. It has been suggested in that experimental report [28] that the thermal stability of the N-terminal $\beta$ hairpin is reflected in the *A* state or the molten globule state (MGS) of the structure of Ubiquitin where the same hairpin found as unaffected. Such kinds of suggestions could be extended to correlate the structure of TS and the *A* state. The TS is a slightly disrupted form of the native structure and further disruption from TS leads to a structure which is very close to the reported MGS of the protein, which is not a stable state in water (as observed from the simulation and experiments) [14,40]. Most of the secondary structural elements are common to both TS and MGS, except the $\beta3/\beta5$ interaction, which is lacking in the MGS. So for ubiquitin the MGS contains most of the topological information of the TS. From the conformational analysis (Fig. 8) there is indication of the possibility of the presence of a transient intermediate state that is structurally close to MGS. The $\beta1/\beta5$ interaction is very crucial for maintaining the 3D scaffold of the protein that lasted very long (the transition from fourth to fifth column in Fig. 9) and its high thermal resistance is very significant. We guess that formation of $\beta$ sheet at this $\beta1/\beta5$ interface can narrow the conformational search during the folding process.

There had been several reports of protein unfolding simulations at similar temperatures conducted mostly by Daggett and co-workers and similar conclusions were made, i.e., increasing the simulation temperature only activates the unfolding process without grossly modifying it [7]. Similar observations were reported for a different protein [7] and our finding helps to generalize the idea for "small globular proteins." The observation that the transition intermediates, despite high structural variability, retain nativelike contacts including some of the secondary structural elements, has been recently suggested to be a common feature in the folding of small globular proteins [10]. As we have mentioned earlier in this article that there was some controversy about the TS of the unfolding process obtained from the experimental results. From our observation we suggest that TS determined from the $\psi$-value analysis should be considered as the correct TS rather than that detected from the $\Phi$-value analysis.

## ACKNOWLEDGMENTS

We thank DIC, Department of Biophysics, University of Calcutta for their computational facility. S.G.D. is thankful to CSIR, India for financial support through CSIR-NET.

[1] C. Levinthal, J. Chem. Phys. **65**, 44 (1968).
[2] T. E. Creighton, *Protein Folding* (W. H. Freeman, New York, 1992).
[3] D. J. Brockwell, D. A. Smith, and S. A. Radford, Curr. Opin. Struct. Biol. **10**, 16 (2000).
[4] N. Ferguson and A. R. Fersht, Curr. Opin. Struct. Biol. **13**, 75 (2003).
[5] V. Daggett, Curr. Opin. Struct. Biol. **10**, 160 (2000).
[6] V. Daggett and A. R. Fersht, Trends Biochem. Sci. **28**, 18 (2003).
[7] R. Day, B. J. Bennion, S. Ham, and V. Daggett, J. Mol. Biol. **322**, 189 (2002).
[8] A. Caflisch and M. Karplus, J. Mol. Biol. **252**, 672 (1995).
[9] B. Rizzuti, V. Daggett, R. Guzzi, and L. Sportelli, Biochemistry (Mosc.) **43**, 15 604 (2004).
[10] K. Lindorff-Larsen, P. Rogen, E. Paci, M. Vendruscolo, and C. M. Dobson, Trends Biochem. Sci. **30**, 13 (2005).
[11] B. Ibara-Molero, V. V. Lodadze, G. I. Makhatadze, and J. M. Sancehez-Ruiz, Biochemistry (Mosc.) **38**, 8138 (1999).
[12] M. Kakuta, A. D. Jayawickrama, A. M. Wplters, A. Manz, and J. V. Sweedler, Anal. Chem. **75**, 956 (2003).
[13] M. M. Harding, D. H. Williams, and D. N. Woolfson, Biochemistry (Mosc.) **30**, 3120 (1991).
[14] B. J. Stockman, A. Euvrard, and T. A. Scahill, J. Biomol. NMR **3**, 285 (1993).
[15] C. G. Benitez-Cardoza, K. Stott, M. Hishberg, H. M. Went, D. N. Woolfson, and S. E. Jackson, Biochemistry (Mosc.) **43**, 5195 (2004).
[16] C. D. Snow, H. Nguyen, V. S. Pande, and M. Gruebele, Nature (London) **420**, 102 (2002).
[17] V. V. Loladze, D. N. Ermolenko, and G. I. Makhatadze, Protein Sci. **10**, 1343 (2001).
[18] D. Roccatano, I. Daidone, M. A. Ceruso, C. Bossa, and A. Di. Nola, Biophys. J. **84**, 1876 (2003).
[19] A. Cavalli, P. Ferrara, and A. Caflisch, Proteins: Struct., Funct., Genet. **47**, 305 (2002).
[20] F. Rao and A. Caflisch, J. Mol. Biol. **342**, 299 (2004).
[21] S. Gnanakaran, H. Nymeyer, J. Portman, K. Y. Sanbonmatsu, and A. E. Garcia, Curr. Opin. Struct. Biol. **13**, 168 (2003).
[22] K. Lindorff-Larsen, M. Vendruscolo, E. Paci, and C. M. Dobson, Nat. Struct. Biol. **11**, 443 (2004).
[23] X. Wu and B. R. Brooks, Biophys. J. **86**, 1946 (2004).
[24] A. Cavalli, U. R. S. Haberthur, E. Paci, and A. Caflisch, Protein Sci. **12**, 1801 (2003).
[25] M. Vendruscolo, N. V. Dokholyan, E. Paci, and M. Karplus, Phys. Rev. E **65**, 061910 (2002).
[26] T. Sivaraman, C. B. Arrington, and A. D. Robertson, Nat. Struct. Biol. **8**, 331 (2001).
[27] M. Jourdan and M. S. Searle, Biochemistry **40**, 10 317 (2001).
[28] S. Grzesiek and F. Cordier, J. Mol. Biol. **715**, 739 (2002).
[29] T. R. Sosnick, R. S. Dothager, and B. A. Krantz, Proc. Natl. Acad. Sci. U.S.A. **101**, 17 377 (2004).
[30] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, W. R. Impey, and M. L. Klein, J. Chem. Phys. **79**, 926 (1983).
[31] A. D. MacKerell *et al.*, J. Phys. Chem. B **102**, 3586 (1998).
[32] L. Harr, J. S. Gallagher, and G. S. Kell, *NBS/NRC Steam Tables* (Hemisphere, Washington, DC, 1984).
[33] G. S. Kell, J. Chem. Eng. Data **12**, 66 (1967).
[34] M. P. Allen and D. J. Tildesley, *Computer Simulation of Molecular Liquid* (Clarendon Press, Oxford, 1987).
[35] S. G. Dastidar and C. Mukhopadhyay, Phys. Rev. E **68**, 021921 (2003).
[36] A. Li and V. Daggett, Proc. Natl. Acad. Sci. U.S.A. **91**, 10 430 (1994).
[37] A. Li and V. Daggett, J. Mol. Biol. **257**, 412 (1996).
[38] M. Levitt, J. Mol. Biol. **168**, 621 (1983).
[39] D. P. Goldenberg and T. E. Creighton, Biopolymers **24**, 167 (1985).
[40] D. O. V. Alonso and V. Daggett, J. Mol. Biol. **247**, 501 (1995).