Available online at www.sciencedirect.com
SCIENCE DIRECT®
Polymer 45 (2004) 581–595
polymer
www.elsevier.com/locate/polymer

Unfolding events of Chymotrypsin Inhibitor 2 (CI2) revealed by Monte Carlo (MC) simulations and their consistency from structure-based analysis of conformations

S. Banu Ozkan[1], Gul Safak Dalgýn[1], Turkan Haliloglu*

Department of Chemical Engineering, Polymer Research Center, Bogazici University, Bebek, Istanbul 80815, Turkey

Received 2 June 2003; received in revised form 7 October 2003; accepted 7 October 2003

Abstract

We investigated the behavior of the conformations of Chymotrypsin Inhibitor (CI2) from the native to the denatured states, obtained in Monte Carlo (MC)/Metropolis simulations, where a low-resolution model is used together with knowledge-based potentials. New conformations starting from the X-ray native structure are generated by random perturbations along with a constraint to increase the radius of gyration. Unfolding is also simulated by unrestrained simulations at a higher temperature. All simulations yield a similar sequence of unfolding events. The preferred pathway starts with loss of native contacts between (N-terminal)-β_3 and continues with β_2–β_3. The persistence of the contacts between β_1 and β_2 at intermediate values of the fraction of native contacts (Q); whereas, highly unfolded conformations with only some helical contacts persisting at low values of Q, are observed. Structure-based analysis of the fluctuations of the unfolded conformations by Gaussian Network Model (GNM) reveals that the termini of the chain—C terminus being more mobile—depict relatively higher flexibility with a native-like hinge near β_2 that divides the structure into two domains. The fluctuations of the two domains are negatively correlated, with partly folded α-helix and a small hydrophobic cluster in the middle of the chain displaying positively correlated fluctuations. The most persistent short-range rotational bond correlations are observed between the residues of α-helix, C terminus of the β_1-part of the reactive site loop, and around the C terminus of the β_2. The latter regions also appear as hot spots; i.e. high frequency fluctuating regions, of the structure surviving in unfolded conformations. The results imply that the unfolded CI2 has an intrinsic ability to undergo correlated fluctuations along with some residual native structure specifically induced by its sequence, consisting at the lowest level of a single hinge.

Keywords: Gaussian Network Model; Fluctuations; Slow/fast modes

1. Introduction

The spontaneous folding of protein molecules with a huge number of degrees of freedom into a unique three-dimensional structure that carries out a biological function is the simplest case of biological self-organization. Thus, there have been many efforts to understand how a protein folds into its native structure, yet, little is known experimentally about the dynamic behavior of unfolded protein. Learning the structural characteristics of the partially folded structures can help us to elucidate how proteins fold. During the protein-folding process, a polypeptide chain can form partially folded intermediate states to avoid searching a large conformation space on the energy landscape. Recent NMR experiments with partially aligned protein molecules in strongly denaturing conditions suggest that the unfolding is less chaotic than is widely believed [1]. It has been suggested that the steric clashes between side chains and backbone lead to the long-range order in the structure [2]. The discussion of a network of hydrophobic clusters, which would explain the long-range order observed in urea-denatured stapylococcal nuclease [3] and water and urea-denatured lysozyme [4], stimulate the reassessment of the accepted concepts about denatured proteins. Protein folding in vivo may be quite different from in vitro, as it may fold co-translationally and protein folding in vivo sometimes needs the assistance of molecular chaperones. Thus,

* Corresponding author. Tel.: +90-212-358-1540; fax: +90-212-257-5032.
E-mail address: turkan@prc.bme.boun.edu.tr (T. Haliloglu).
[1] Both authors equally contributed.

0032-3861/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/j.polymer.2003.10.092

studying the structural dynamics of the denatured protein can help us learn about the physics and chemistry principles that govern the protein folding.

Since Anfinsen's [5] original demonstration of spontaneous protein refolding, major advances have occurred in experimental studies to determine the mechanism of folding [6–8]. There have also been many theoretical approaches that shed light on the nature of folding pathways, their transition states (TS), and the role of intermediates in folding. One of the significant progresses in theoretical approaches was the energy landscape theory [9,10] that gave us the general framework of the folding mechanisms based on statistical physics. So far the most important insights have come from simulations on simplified representations of proteins, using on-lattice and off-lattice models [10–15]. Such models provide simple examples that can help to clarify basic principles of folding kinetics. On the other hand, all-atom molecular dynamic (MD) simulations can give the most detailed and realistic information. Nevertheless, these simulations have been restricted by computer time to one or just a few trajectories of tens of nanoseconds. To the extent that early unfolding steps could be considered as analogous to the late folding steps when considered in the reverse direction, MD unfolding simulations can provide useful information about the key interactions that should play a role in forming the native structure. Since the specific interactions between the residues and solvent play an important role in the stability of the native structure, it is useful to carry out such simulations at atomistic detail. This comes with the problem of timescale of folding/unfolding that is several order of magnitude larger than those currently attainable by MD simulations. Thus, it is necessary to carry out simulations at higher temperatures than room temperature to destabilize the native structure, or with the use of driving potentials that promote unfolding. In the former case, large entropic contributions may unrealistically destabilize some parts of the structure before the others, or may alter the mechanism by which the protein folds. In the latter case, where driving potentials are used, it is important to be cautious that no significant biases are introduced that might favor a particular unfolding pathway; this can be checked in principle by comparing the chronology of unfolding conformational events in separate unrelated simulations, and comparing them to those observed experimentally. The most commonly used driving potentials make use of radius of gyration [16,17] and the root-mean-square deviation [18,19] from the native structure.

The all-atom MD simulations of unfolding trajectories of Chymotrypsin Inhibitor 2 (CI2) [20,21] at high temperatures gave insights into the nature of the TS and denatured state ensemble. Lazaridis and Karplus [22] also analyzed the unfolding MD trajectories of CI2 to clarify the TS. They observed that the TS region for folding and unfolding occurs early with only 25% of the native contacts. Another interesting result from the latter simulations was the emergence of the statistically preferred unfolding pathway. Zhou and Karplus [23] used a discrete MD technique to study the folding of the small three-helix bundle fragment of Protein A. In this study, the two different dominant trajectories (fast and slow tracks) were observed for the folding of helical proteins when the single energy parameter (the difference between the strength of native and non-native contacts) was changed. Dokholayan et al. [24] investigated the protein folding nucleus of 46-mer off-lattice protein model using MD. The simulations revealed that a few well-defined contacts were formed with high probability in the TS that drives the protein into its folded conformation.

Monte Carlo (MC) dynamic simulation methods, on the other hand, have also been widely used to understand the basic folding principles of on-lattice simple protein model chains [25–28]. The statistical analysis of hundreds of folding MC trajectories of lattice model proteins performed by Pande and Rokhsar [27], revealed a classical dominant folding pathway which was composed of on-pathway intermediates. Klimov and Thirumalai [29] investigated the TS structure and the folding nuclei of 27-mer and 36-mer lattice models. The analysis of individual trajectories showed that the polypeptide chain reaches the native state upon the formation of critical contacts, which is consistent with the nucleation-collapse mechanism. The effect of non-native contacts on the folding mechanism was analyzed by Li et al. [25] using MC techniques. They also investigated the TS structure of CI2 with all atom model simulations and found that the α-disrupted states have a stronger tendency for folding than the β-disrupted states [30].

The idea of studying the small, single protein, CI2 was a turning point, which gave insight into protein folding and new experimental techniques. CI2 was found to fold and unfold as a simple two-state system with no kinetic intermediates. Mutational analysis, Φ-value analysis [31,32], allows the investigation of the role of individual residues in the TS for folding. This method has also been used by many theoreticians to understand the TS structure and the folding mechanism of simple protein models [22,25,29,33,34]. This method has stimulated interactions between experimentalists and theoreticians [35] as it allows understanding the folding mechanisms.

Here, we aim to analyze the unfolding conformations of CI2 obtained by adopting a MC simulation technique on a low-resolution protein model, which has been previously applied [36–38] for the native state dynamics of proteins, for the unfolding simulations with a restraint on the radius of gyration, S_g, that promotes unfolding. The simulations are also carried out without restraint but at a higher temperature to provide a basis of comparison. The unfolding trajectories of CI2 by 10 independent simulations for unrestrained and restrained simulations are analyzed, with emphasis on the results from the restrained simulations that let us carry out a systematic analysis on the unfolded structures with higher volume. As will be elaborated in Section 3, there is a

statistically preferred pathway observed for the unfolding of CI2 that is consistent with experimental observations and MD simulations. Structural-based analysis carried out for the conformations extracted for the trajectories from the native to denatured states with the Gaussian Network Model (GNM) allow us to study the cooperative motions in terms of the cross-correlations of the residue fluctuations and to search for the slowest/fastest modes of dynamics as different states on the unfolding pathway are visited.

2. Model and method

2.1. Native structure

The simulations begin with the X-ray native structure [39] (PDB [40] code: 2ci2). CI2 is a 64-residue protein that consists of an α-helix and a three-strand β-sheet (Fig. 1). The main hydrophobic core is formed by the packing of the α-helix against the β-sheet. It was the first protein shown to fold by a two-state mechanism, involving only the denatured and native states, which are separated by the energetically unfavorable TS [41].

2.2. Low resolution model

In the present study, each residue i is represented by two interaction sites; its α-carbon atom (C_i^{α}) and its side-chain interaction center (S_i). The sites on the side-chains are selected on the basis of the specific structure and energy characteristics of the amino acids [42]. The peptide backbone is represented by the virtual bond model, originally proposed by Flory and collaborators [43]. A peptide of N residues has $N-1$ virtual bonds connecting successive α-carbons. Virtual bonds are highly stiff and were taken here as fixed at their equilibrium values, l_i, (i.e. the bond between residues i and $(i-1)$) all having length 3.81 ± 0.03 Å. Thus, the peptide backbone conformation can be defined by the $2N-5$ dimensional vector $\{\theta_2, \theta_3, \ldots, \theta_{N-1}, \phi_3, \phi_4, \ldots, \phi_{N-1}\}$ corresponding to $N-2$ virtual bond angles (θ_i) at the i-th carbon, and $N-3$ dihedral angles (ϕ_i) at the i-th virtual bond. In the same way, the distance between S_i and C_i^{α} (l_i^s) and the angle between l_i and l_i^s (θ_i^s) are fixed at their equilibrium values. The conformation of side-chain i can be expressed by the torsion angle (ϕ_i^s) that is defined by l_{i-1}, l_i, and l_i^s.

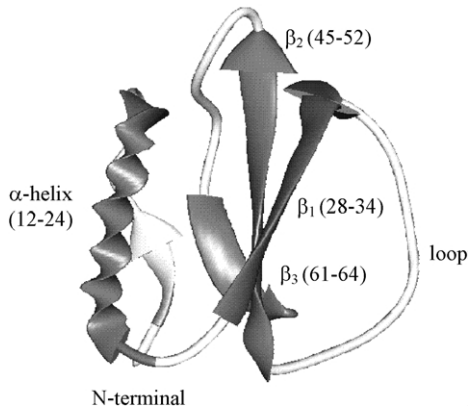

Fig. 1. Ribbon representation of the main-chain fold of CI2. The diagram is based on the crystal structure with PDB code 2CI2 determined to 2 Å resolution [3]. The residues were renumbered to begin with 1 instead of 20 for ease of comparison with previous studies. The residues forming the secondary structure elements are indicated in parentheses according to the numbering scheme.

The energy of the conformation Φ is found from additive contributions of short-range interaction potentials $E_{SR}\{\Phi\}$, between covalently bonded units along the sequence, and long-range interaction potentials $E_{LR}\{\Phi\}$, between non-bonded residues that are close in space.

$$E\{\Phi\} = E_{SR}\{\Phi\} + E_{LR}\{\Phi\} \tag{1}$$

Here $E_{LR}\{\Phi\}$ [44] is calculated from the following expression:

$$E_{LR}\{\Phi\} = \sum_{i=1}^{N-3}\sum_{j=i+3}^{N} W_{SS}(r_{ij}) + \sum_{i=1}^{N-4}\sum_{j=i+4}^{N} W_{SB}(r_{ij}) + \sum_{i=1}^{N-5}\sum_{j=i+5}^{N} W_{BB}(r_{ij}) \tag{2}$$

where r_{ij} is the distance between the sites i and j in conformation $\{\Phi\}$. $W_{SS}(r_{ij})$ is the potential between side-chains i and j, $W_{SB}(r_{ij})$ is the potential between the side-chain and backbone sites of the ith and jth residues, and $W_{BB}(r_{ij})$ is the potential between two backbone interaction sites i and j. Here, the interacting sites are separated by at least five intervening virtual bonds.

The short-range conformational energy is calculated as described by Bahar and collaborators [45] using the formulation

$$E_{SR}(\Phi) = \sum_{i=2}^{N-1} E(\theta_i) + \sum_{i=2}^{N-1} [(1/2)(E(\phi_i^-) + E(\phi_i^+)) + \Delta E(\phi_i^-, \phi_i^+)] + \sum_{i=2}^{N-1} [\Delta E(\theta_i, \phi_i^-) + \Delta E(\theta_i, \phi_i^+)] \tag{3}$$

The first term is to account for distortion of bond angles and the second is for the bond torsions, in which ϕ_i^- and ϕ_i^+ refer to the rotational angles of the virtual bonds preceding and succeeding the ith α-carbon, respectively. The other terms account for pairwise interdependence of the torsion and/or bond angle distortions.

A MC/Metropolis simulation scheme is used to generate a series of low-resolution conformations from native state to highly extended conformations by applying random perturbations to generalized coordinates of the structure. The attempted perturbation is accepted or rejected according to

the Metropolis criterion as

$$\exp\{-(E\{\Phi\}-E\{\Phi\}_0)/RT\} \leq r \quad (4)$$

Here, $E\{\Phi\}_0$ and $E\{\Phi\}$ are energies of the conformations before and after perturbation, respectively, using the potentials described in Eqs. (1–3). One MC step is composed of N (number of residues = 64) number of random perturbations with maximum variation of δk to the generalized coordinates of ϕ_i and θ_i, and ϕ_{is}. δk is 20° for ϕ_i and 5° for θ_i and ϕ_{is}. In restrained simulations (at $2kT$), an additional criterion comes with the constraint, which promotes an increase in the radius of gyration of each generated conformation as a function of the MC step, computed as $m \times$ number of MC step $\times$ radius of gyration of the native conformation, where $m = 0.0005$. Thus, the radius of gyration of a generated conformation at the 10,000 MC steps would be 5 times higher than that of the native conformation. Five independent simulations, each having 15,000 MC steps are carried out. Simulations are run for various values of m to test how it affects the unfolding behavior. To have a basis for comparison, five independent runs of unrestrained simulations but at a higher temperature ($4kT$) are also carried out. In the latter simulations, the only condition is the radius of gyration of the generated conformations not be lower than that of the native state to avoid more collapsed conformations.

In the course of this study, in addition to the two types of simulations above, additional simulations are carried out. In these simulations, the extent of the potential of interactions governing the unfolding behavior is modified. Incorporating only short-or long-range interactions, reducing the barrier heights of the transitions by hyperbolic sampling of the energy surface and considering only first coordination shells for the interactions at different temperatures are considered. These simulations allowed us to search how robust the methodology is and the effect of imposed changes on the observed unfolding behavior.

2.3. Gaussian Network Model (GNM)

Recently, a simple analytical model was developed to describe the vibrational dynamics of proteins, referred to as the Gaussian Network Model (GNM) [46]. GNM uses the known topology of protein-protein contacts to model the protein as an elastic network with uniform single-parameter harmonic potentials between close-lying α-carbons. The equilibrium correlation between the fluctuations $\Delta\mathbf{R}_i$ and $\Delta\mathbf{R}_j$ of residues i and j is given [46] by

$$\langle \Delta\mathbf{R}_i \cdot \Delta\mathbf{R}_j \rangle = (3k_B T/\gamma)[\mathbf{\Gamma}^{-1}]_{ij} \quad (5)$$

where $\mathbf{\Gamma}$ is a symmetric matrix known as the Kirchoff or connectivity matrix of α-carbons atoms including all residue pairs within a first interaction shell (r-cutoff = 7.0 Å) [44]. The favorable agreement between the residue fluctuations of X-ray crystallographic temperature factors (B-factors) and those obtained from Kirchoff matrix indicates that the cutoff distance of 7.0 Å is good enough to analyze the vibrational dynamics of the present system. $\mathbf{R}_i$ is the position vector of the i^{th} α-carbon, k_B is the Boltzmann constant and T is the absolute temperature. γ is the normalization constant that is the counterpart of the single parameter of the Hookean pairwise potential originally proposed by Tirion [47].

The inverse of $\mathbf{\Gamma}$ can be written as

$$\mathbf{\Gamma}^{-1} = \mathbf{U}(\mathbf{\Lambda}^{-1})\mathbf{U}^T \quad (6)$$

where $\mathbf{U}$ is an orthogonal matrix whose columns $\mathbf{u}_i$ are the eigenvectors of $\mathbf{\Gamma}$, and $\mathbf{\Lambda}$ is the diagonal matrix of the eigenvalues (λ_i) of $\mathbf{\Gamma}$. Mean-square fluctuations of the α-carbon atoms and the cross correlations between them are given by the respective diagonal and off-diagonal elements of $\mathbf{\Gamma}^{-1}$.

It is possible to decompose $\mathbf{\Gamma}^{-1}$ into the sum of contributions from individual modes as

$$\mathbf{\Gamma}^{-1} = \sum_{k=2}^{N} \lambda_k^{-1}\mathbf{u}_k\mathbf{u}_k^T = \sum_{k=2}^{N} \mathbf{A}^{(k)} \quad (7)$$

Here $\mathbf{A}^{(k)}$ is the $N \times N$ matrix (for a protein of N residues) describing the contribution of the k^{th} vibrational mode to atomic fluctuations. The first eigenvalue of $\mathbf{\Gamma}$, identically equal to zero, is not included in the above summation. The i^{th} eigenvalue represents the frequency of the i^{th} mode of motion, and the i^{th} eigenmode gives the shape of this mode as a function of the residue index.

Thus, the dynamics of a bimolecular system can be decomposed by GNM into a collection of internal modes of different frequencies similar to normal mode analysis [48], and GNM provides a detailed description of the dynamics around a local energy minimum with its inherent limitations, yet it provides much insight into protein dynamics. The fast modes reflect the localized motions involving high-frequency fluctuations of individual residues [49]. These residues are generally in dense regions of the structure with high coordination numbers, and they are potentially important for the stability. The slowest modes, with the lowest frequencies, on the other hand, refer to the most cooperative motions involving the overall structure, which give information about the MDs relevant to biological function occurring on the global scale [48–53]. GNM has been applied to a number of different biomolecular systems including RNA complexes [50], enzyme complexes [51], substrate-binding proteins [52] and monomeric proteins [53], and it has been shown to compare favorable with X-ray crystallographic temperature factors [46], H/D exchange behavior [54] and order parameters from NMR-relaxation measurements [55].

A similar approach to GNM was recently introduced by Micheletti et al. [56]. The self-consistent pairwise contact probabilities were used in the latter method to describe the interactions between the pairs as a function of temperature.

In the limit, where $T = 0$, the method is the same as GNM; and upon increasing the temperature, the decrease in the pairwise contact probabilities yields milder interactions between the pairs. The analysis of the different strengths of the interactions between the interacting pairs from the lowest to the higher temperatures showed that the slowest relaxation time, which dominates the long time relaxation kinetics, correlates with the experimentally determined folding rates. Shen et al. [57] expanded the model of Michelleti [58] by including an external force field and studied the force induced unfolding of a helical segment of transformylase and beta-stranded domains of titin. All these studies show that the near-equilibrium dynamics may provide insights into the folding processes.

In the present analysis, the GNM method is used to carry out structure-based analyses of the conformations extracted from the trajectories at different stages of the simulations, given that each conformation with its topology represents one of the minimum states on the energy surface that characterizes the unfolding. This enables us to see the potential dynamics behavior of the states visited during the unfolding. For this, the two extreme ends of the spectrum, the slowest and the fastest dynamics modes of the residue fluctuations are analyzed and the correlations between the fluctuations are searched as the number of native contacts diminishes.

3. Results and discussion

3.1. Radius of gyration, S_g, and root-mean-square deviation, RMSD, of generated conformations from native to denatured state

Fig. 2 presents the square of radius of gyration, S_g, and the root-mean square deviation, RMSD, of the generated conformations from the native structure versus MC time steps for one representative run for both constrained and unrestrained simulations. The radius of gyration increases linearly as expected in the restrained simulation; whereas, it fluctuates in the course of the simulation in the unrestrained simulation. The S_g of the denatured state was observed as $\sim$80–100% larger than that of the native state in MD simulations [21] of CI2; that is the value we observe at around $\sim$4000 MC time steps. On the other hand, RMSD values follow an increasing trend continuously until it stabilizes around 6,000 MC steps in constrained simulations; yet, it depicts a sharp increase until about 500 MC steps and then reaches a plateau in unrestrained simulations. RMSD values from both simulations seem to overlap at around 6,500 MC steps. The imposed restraint and lower temperature in the restrained simulations delay the equilibration of RMSD values compared to the unrestrained simulations, nevertheless allow us to carry out systematic/controllable analysis of the conformations on the unfolding pathway.

3.2. Energetics

The total energy per residue, ($\Delta E/nkT$), based on the knowledge based-potentials [44,45] described above, is computed and plotted as a function of MC time steps in Fig. 3 for one representative run from both restrained and unrestrained simulations. The energy seems to approach equilibrium at relatively later times in the restrained simulations compared to the behavior of the energy in the unrestrained simulations. This is mainly because of the lower temperature and imposed restraint on S_g that lowers the acceptance ratio of the moves in the restrained simulations. Nevertheless, it is possible to observe the convergence of both curves at later stages of the simulation (not shown in the figure). The analysis of the constituents of the total energy reveals that the long-range interactions contribute to the unfolding more compared to the short range interactions and the contributions of the constituents of the long-range interactions rank from the highest to lowest as: backbone–backbone > side-chain–side-chain > side-chain–backbone interactions in both simulations.

Regarding the relative contributions of the short-range and long-range interactions to the unfolding, the simulations performed including only long-range or short-range interactions also reveal the importance of the long-range interactions. Although omission of the short-range interaction does not affect the observed sequence of events and the features of the unfolding, a rapid unfolding with unrecognizable pattern is observed in the absence of long-range interactions. On the other hand, reducing the barrier heights by the hyperbolic sampling of the energy surface does not change the observed events of the unfolding, whereas including only the first coordination shell of interactions leads to a rapid unfolding without any pattern.

3.3. Unfolding trajectory

The fraction of the native contacts, Q, which is the ratio of the number of native contacts in each state to the total number of native contacts in the native state, is computed. Two residues are considered to be in contact, excluding the first and the second near neighbors, if the distance between their α-carbons is less than or equal to 7 Å (first coordination shell) [42,44,45]. To elaborate the loss of the native contacts in the course of unfolding simulations, the contact maps of the generated conformations in different snapshots are constructed for all 10 simulations. The contact maps of the conformations throughout the unfolding display similar characteristics in all simulations performed at different temperatures with and without radius of gyration restraint. The recent MD analysis of CI2 at different elevated temperatures [59] also shows that overall characteristics and sequence in the unfolding process are well conserved at all temperatures despite the fact that the unfolding accelerates with increasing temperature. The

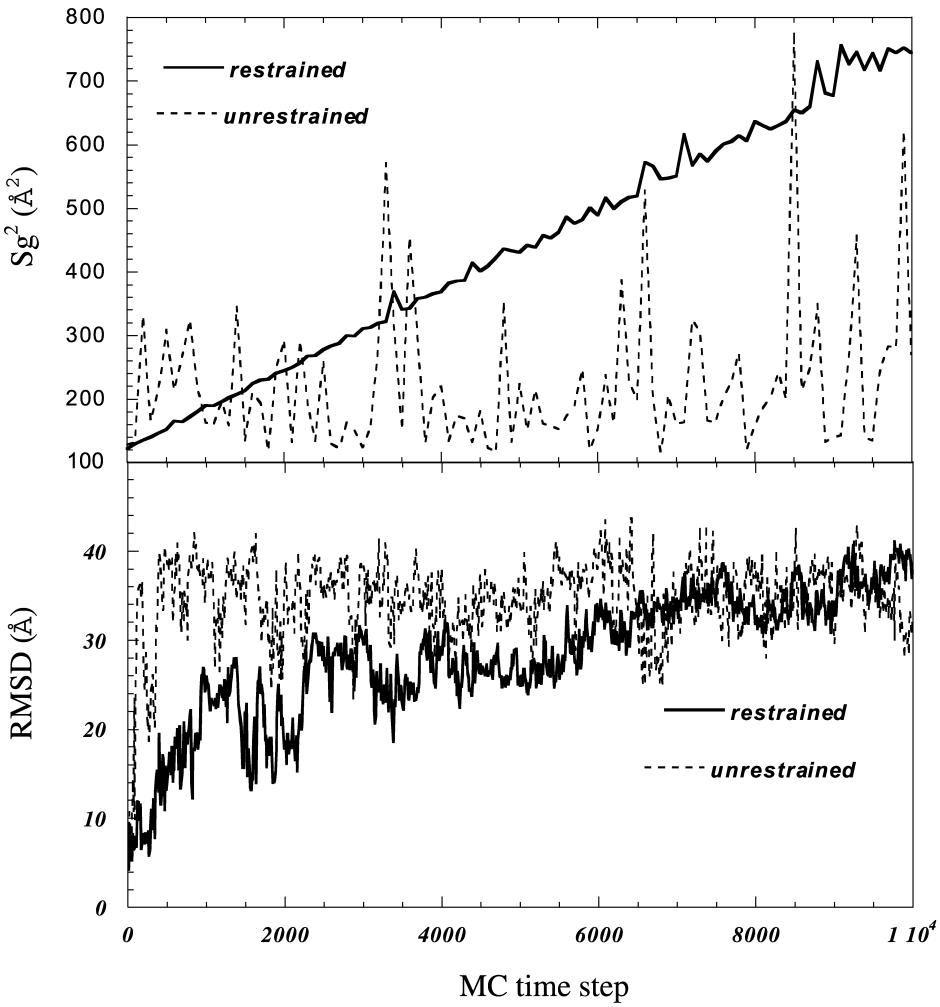

Fig. 2. The radius of gyration squared, S_g^2, (a) and the root-mean-square deviation, RMSD, (b) from the native state as a function of MC time step. One representative run is chosen from restrained simulations where the radius of gyrations linearly increases with MC steps at $2kT$ and unrestrained simulations at a higher temperature of $4kT$.

analysis of the ten simulations reveals that there is a statistically preferred path that shows a sequence of events for the unfolding of CI2, consistent with the previous MD studies [18,21,22,30]. Fig. 4 presents the contact map of the native state conformation (a), the conformations with $Q = 0.5$ (b), 0.35 (c) and 0.1 (d) from one of the restrained simulation runs. The preferred pathway starts with the disappearance of native contacts between N-terminal and β_3. MD simulation study by Kazmirski et al. [21] also shows that the loop between the β-strands 2 and 3 moves away from the helix to expose the hydrophobic core. After the exposure of the hydrophobic core, the contacts between the residues in β_2 and β_3 are lost. The loss of persistent contacts between residues in β_1 and β_2 accelerates the complete unfolding. The contacts near α-helix and the α-helix$-\beta_1$ loop may appear/disappear at later stages of the simulation. In general, the conformations in TS and denatured state display features that are similar to the observations in the MD simulations by Kazmirski and coworkers [21] and from TMD (Targeted MD) simulations by Ferrara and coworkers [18], in which a restraint promotes either the increase of the conformational distance from the native state or the decrease of the distance from a target unfolded structure. In the latter study, as a basis of comparison, unfolding is also simulated by conventional, i.e. unrestrained MD simulations where the sequence of unfolding events of CI2 is found to be similar in all simulations.

3.4. Persistent pair of residues

The interacting residue pairs (contact pairs) that are observed in all simulations from the conformations with $0.5 > Q > 0.25$ (a) and $0.25 > Q > 0.1$ (b), with their total number of contacts (i.e. total number of contacts that a residue has in the native state) are given in Tables 1 and 2, respectively. The analysis of the conformations in the ensemble (a) yields that the most persistent contacts are mainly between the residues within the α-helix and between the residues of β_1 and β_2. Of these residue pairs, the interactions between Ser12-Glu15, Ser12-Ala16,

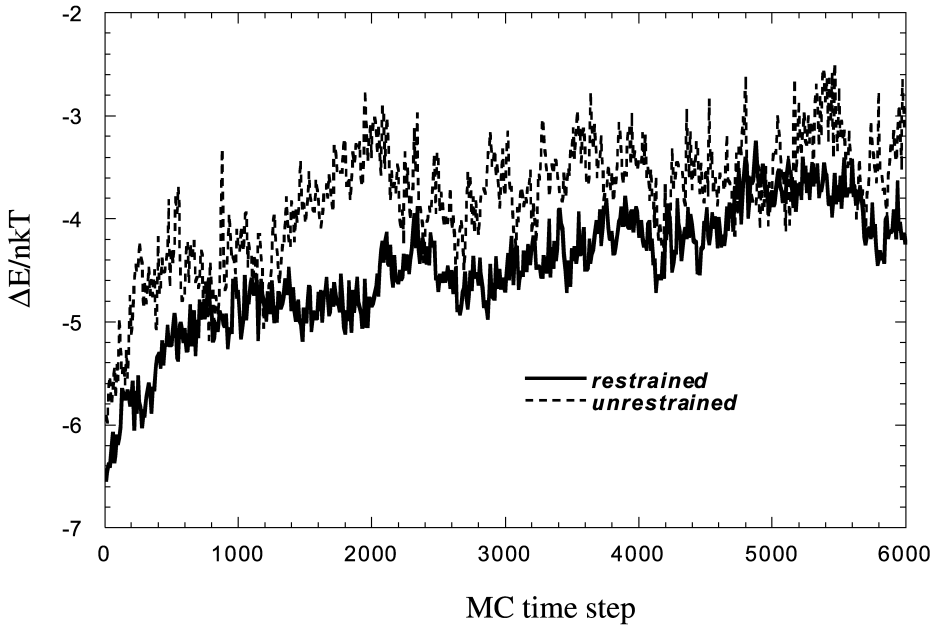

Fig. 3. The total energy versus MC time step for one of the representative runs from restrained and unrestrained simulations.

Val13-Ala16, Glu14-Lys17, Lys17-Ile20 and Lys17-Leu21 still exist in the ensemble conformations with very low Q (b). In the latter ensemble, in particular towards the lower range of Q, there also emerge some pairs of non-native contacts around both termini. This is mainly because of the enhanced mobility of termini, which promotes the occurrence of some possible contacts nearby, in highly denatured conformations. The persistence of these non-native contacts may indicate their significance in folding of CI2. The formation of non-native contacts during the folding process

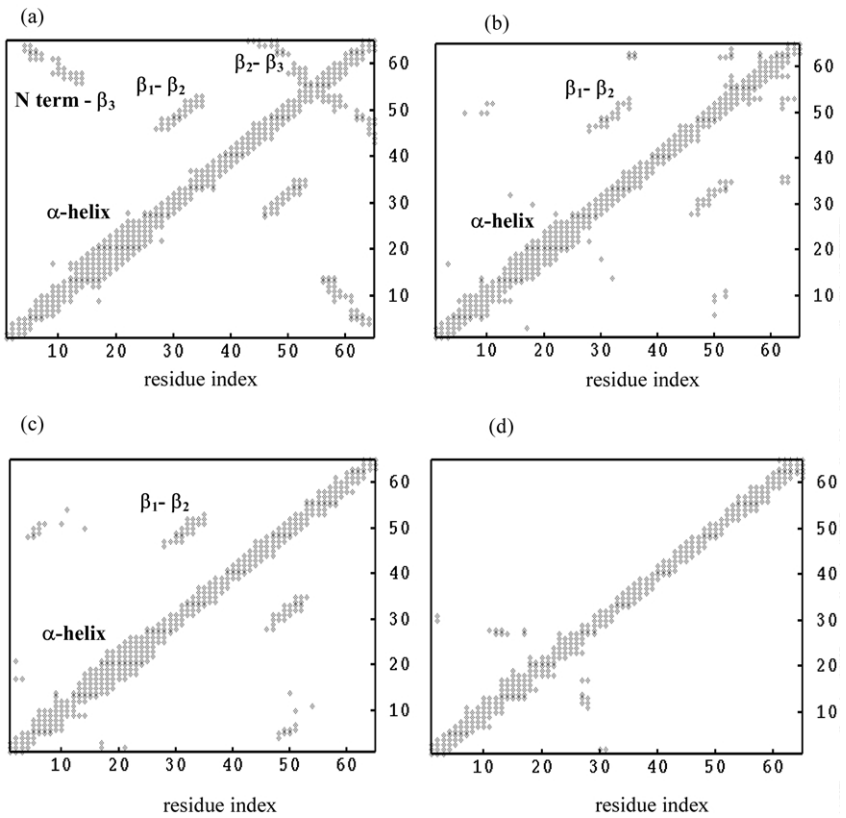

Fig. 4. The contact maps of (a) the native conformation, (b) the conformation at 500 MC time step ($Q = 0.5$), (c) the conformation at 1000 MC time step ($Q = 0.35$), (d) the conformation at 10,000 MC time step ($Q = 0.1$) for one of the restrained runs.

Table 1
The pair of residues that make the most persistent contacts over all unfolding process and the corresponding number of contacts that each residue makes ($0.5 > Q > 0.25$)

Residue pair	Location	Residue names and their numbers of contacts in parentheses	
Leu8-Lys11	(N-term)	Leu8 (9)	Lys11 (9)
Lys11-Glu15	(N-term and α-helix)	Lys11 (9)	Glu15 (8)
Lys11-Ala16	(N-term and α-helix)	Lys11 (9)	Ala16 (10)
Glu15-Lys18	(α-helix)	Glu15 (8)	Lys18 (8)
Ala16-Val19	(α-helix)	Ala16 (10)	Val19 (8)
Ala16-Ile20	(α-helix)	Ala16 (10)	Ile20 (8)
Lys18-Leu21	(α-helix)	Lys18 (8)	Leu21 (9)
Lys18-Gln22	(α-helix)	Lys18 (8)	Gln22 (7)
Val19-Gln22	(α-helix)	Val19 (8)	Gln22 (7)
Val19-Asp23	(α-helix)	Val19 (8)	Asp23 (6)
Ile20-Asp23	(α-helix)	Ile20 (8)	Asp23 (6)
Ile20-Lys24	(α-helix)	Ile20 (8)	Lys24 (7)
Leu21-Lys24	(α-helix)	Leu21 (9)	Lys24 (7)
Leu21-Pro25	(α-helix and α-helix−β_1)	Leu21 (9)	Pro25 (6)
Gln22-Pro25	(α-helix and α-helix−β_1)	Gln22 (7)	Pro25 (6)
Ile29-Val47	(β_1 and β_2)	Ile29 (7)	Val47 (12)
Ile29-Arg48	(β_1 and β_2)	Ile29 (7)	Arg48 (9)
Ile30-Val47	(β_1 and β_2)	Ile30 (8)	Val47 (12)
Ile30-Arg48	(β_1 and β_2)	Ile30 (8)	Arg48 (9)
Ile30-Leu49	(β_1 and β_2)	Ile30 (8)	Leu49 (9)
Val31-Arg48	(β_1 and β_2)	Val31 (7)	Arg48 (9)
Val31-Leu49	(β_1 and β_2)	Val31 (7)	Leu49 (9)
Leu32-Leu49	(β_1 and β_2)	Leu32 (8)	Leu49 (9)
Leu32-Phe50	(β_1 and β_2)	Leu32 (8)	Phe50 (10)
Pro33-Phe50	(β_1 and β_2)	Pro33 (7)	Phe50 (10)
Phe34-Phe50	(β_1 and β_2)	Val34 (7)	Phe50 (10)
Val34-Val51	(β_1 and β_2)	Val34 (7)	Val51 (12)
Arg43-Arg46	(β_1−β_2 loop and β_2)	Arg43 (6)	Arg46 (10)
Val51-Asp55	(β_2 and β_2−β_3 loop)	Val51 (12)	Asp55 (8)
Asp52-Asp55	(β_2 and β_2−β_3 loop)	Asp52 (9)	Asp55 (8)

Table 2
Pairs of residues with the most persistent contacts and their numbers of contacts ($0.25 > Q > 0.1$)

Residue pair	Location	Residue names and their numbers of contacts in parentheses	
*Leu1-Glu4	(N-term)	Leu1 (2)	Glu4 (8)
*Lys2-Trp5	(N-term)	Lys2 (3)	Trp5 (5)
*Thr3-Pro6	(N-term)	Thr3 (5)	Pro6 (6)
*Trp4-Glu7	(N-term)	Trp4 (8)	Glu7 (4)
Pro6-Val9	(N-term)	Pro6 (7)	Val9 (9)
*Pro6-Gly10	(N-term)	Pro6 (7)	Gly10 (7)
Ser12-Glu15	(α-helix)	Ser12 (9)	Glu15 (8)
Ser12-Ala16	(α-helix)	Ser12 (9)	Ala 16 (10)
Val13-Ala16	(α-helix)	Val13 (9)	Ala16 (10)
Glu14-617	(α-helix)	Glu14 (6)	Lys17 (8)
Glu14-Lys18	(α-helix)	Glu14 (6)	Lys18 (8)
Lys17-Ile20	(α-helix)	Lys17 (8)	Ile20 (8)
Lys17-Leu21	(α-helix)	Lys17 (8)	Leu21 (9)
*Asn56-Glu59	(β_2−β_3 loop)	Asn56 (10)	Glu59 (6)
*Val60-Val63	(β_2−β_3 loop and β_3)	Val60 (8)	Val63 (8)
*Pro61-Gly64	(β_3)	Pro61 (9)	Gly64 (7)

Asterisks denote non-native contacts.

has been observed in experimental and theoretical studies [60,61]. In a recent NMR study [21] of CI2, higher mobility of termini with respect to the rest of the chain was also observed. In the simulations of Kazmirski and coworkers [21], the Ile29-Ile37 region was found to make a number of hydrophobic contacts within itself as well as with other hydrophobic regions in the molecule. At intermediate values of Q, the residues Ile29 to Val34 mostly make contacts with Val47 to Val51 of the β_2 region that is part of the main hydrophobic core of the molecule. The correlations between the fluctuations of residues Ile29-Ile37, on the other hand, is observed by structure-based analysis of these conformations, presented below. In a study by Daggett and coworkers [20], where Φ values were calculated to model the TS, the residues Val47 and Leu49 that are in the center of the β-sheet were found to have high Φ-values indicating that interactions made by these residues are retained in the TS. It is also interesting to note that the residues listed in Tables 1 and 2 have high total numbers of contacts, i.e. they have either 6 or more contacts.

3.5. Comparison between experimental and theoretical phi values

Abrupt changes in the evolution of Q can be used to define the candidate structures that form the TS ensemble [18,62]. To this aim, we plot the time evolution of Q as a function of simulation time for all simulations (See Fig. 5(a)). The plots reveal a sharp decrease in Q ($\sim$0.45) at 400–500 MC steps. After that, the structure continuously unfolds and becomes highly unfolded specifically after 3,000 MC steps with Q fluctuating around 0.2–0.1 for a long time and then decreasing to $\sim$0 at much later times in the simulation. We observed the common characteristics of the structures having $Q \sim 0.45$ in which the loop between the β-strands 2 and 3 and the helix are further separated. Consequently, the hydrophobic core is exposed, which, in turn, promotes the unfolding of the whole structure.

To assess the reliability of these candidate TS structures, theoretical Φ-values (Φ_i^{sim}) are compared with the experimental values (Φ_i^{exp}). Following the previous work of Li and Daggett [63], Φ_i^{theo} are calculated as

$$\Phi_i^{\mathrm{sim}} = n_i^{\mathrm{TSC}}/n_i^{N} \tag{9}$$

where n_i^{TSC} is the number of contacts residue i in TS has and n_i^{N} is the number of contacts of the native crystal structure. The number of contacts is calculated by considering α-carbon atom (C_i^{α}) of each residue and its side-chain interaction center (S_i) (i.e. the low resolution model) and contacts are considered to exist if the distance between two non-adjacent C_i^{α} or S_i is less than 5 Å. Thus, Φ_i^{sim} are computed over eight structures that are obtained between 400 and 500 MC steps of the simulation, having Q values

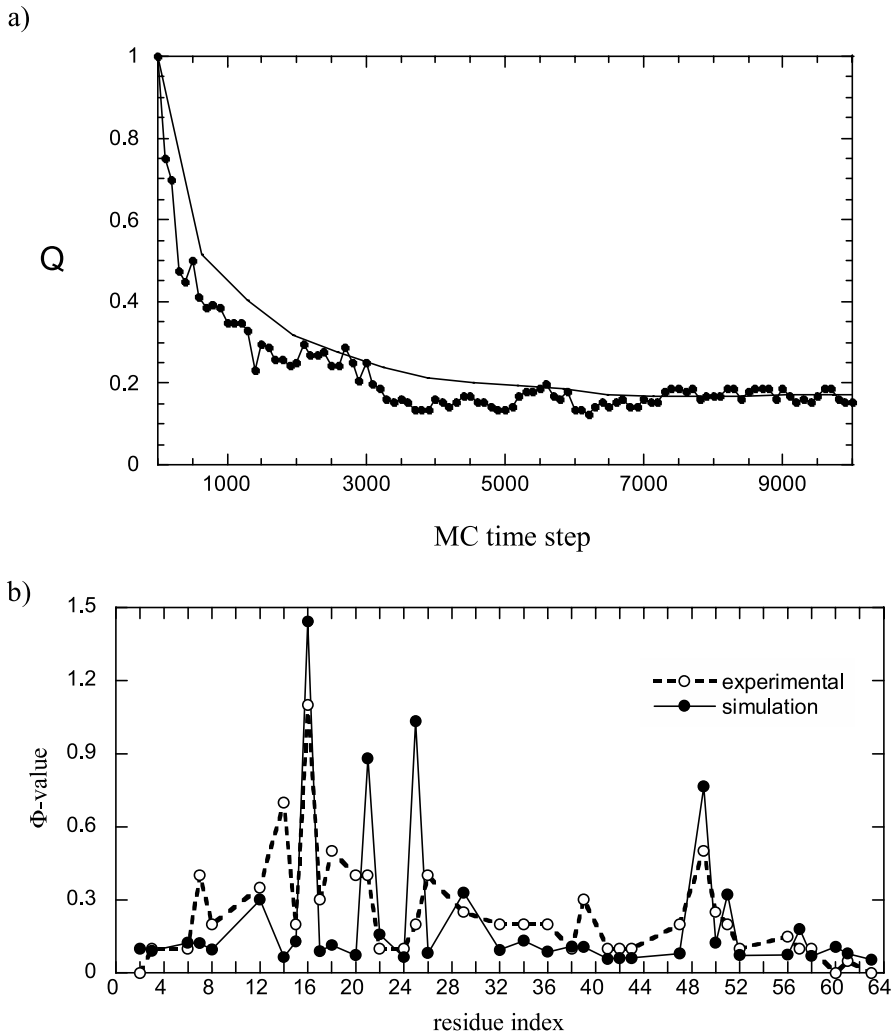

Fig. 5. (a) The ratio of the number of native contacts in each state to the total number of native contacts in the native state, Q, versus the simulation time in MC time steps. (b) Comparison of Φ-values obtained from simulations with the experimentally measured ones [65].

between 0.4 and 0.5. The RMSD between the structures of TS candidates ranges between 3.2 and 5.6 Å.

Fig. 5(b) presents comparison of the calculated Φ-values for TS identified from MC simulations (Φ_i^{sim}) with the experimentally measured Φ-values (Φ_i^{exp}) [64]. The Pearson linear correlation coefficient (ρ) between 37 Φ_i^{exp} values and the corresponding calculated values averaged over eight TS candidate structures is 0.68. In general, there is a good agreement between the calculated Φ-values and experimental ones. Only the residues Glu-14, Leu-21 and Pro-25, show major deviations from the Φ_i^{exp} values. Leaving out these residues yields a ρ of 0.80.

3.6. Cross-correlations between bond rotations

As unfolding is a non-equilibrium process, the unfolding trajectories were separated into three discrete time windows to search for the changes of the rotational angles in a representative time interval τ and the correlation between all pairs of bonds in each window as

$$C_{ij}(\tau) = \frac{\langle \Delta\phi_i(\tau)\Delta\phi_j(\tau)\rangle}{(\langle \Delta\phi_i^2(\tau)\rangle^{1/2}\langle \Delta\phi_j^2(\tau)\rangle^{1/2})} \tag{10}$$

where $\Delta\phi_i(\tau)$ and $\Delta\phi_j(\tau)$ are the changes in the rotational angles of bond i and j within time interval τ, which is taken as 30 MC steps. The bracket indicates the ensemble averages carried out in given time windows. These windows are defined in terms of Q; where (i) $Q > 0.5$, (ii) $0.5 > Q > 0.3$, and (iii) $0.3 > Q$. Fig. 6 presents the negatively and positively correlated pairs. The correlations between the bonds of the α-helix and the loop β_2–β_3 when $0.5 > Q > 0.3$ also evidence the crucial event in the unfolding where the α-helix becomes apart from the β-strands. This result is consistent with the analysis of disappearance of native contacts in Fig. 4 where the unfolding starts with the increase in distance between the α-helix and the loop of β_1 and β_2.

The highest correlations at all stages are observed to be

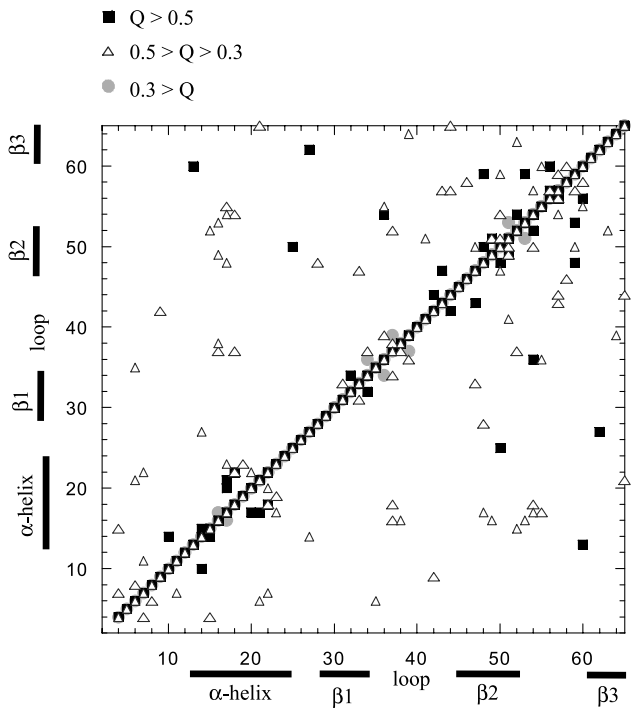

Fig. 6. The correlated (positively and negatively) pairs of bonds having correlation values of C_{ij} ($\tau = 30$ MC time steps) above 0.2 or below -0.2 from the analysis of the conformations in the ranges $Q > 0.5$ (square), $0.5 > Q > 0.3$ (triangle), (iii) $0.3 > Q$ (circle).

in: (i) α-helical part, (ii) C terminus of β_1 and part of reactive site loop and (iii) C terminus of β_2 and loop $\beta_2-\beta_3$. Lys-53 and Leu-54, which are the most conserved residues along the family (alignment data by CLUSTALW which is not shown), appear correlated with some other residues in β_2 at different stages of unfolding. A native-like hinge around the residues at the C terminus of β_2 in the denatured state is also observed by the vibrational dynamic analysis of the denatured conformations by GNM, as will be presented below. On the other hand, Kurt and Haliloglu [37] found the highest correlation between the pair Tyr-42 and Ile-44 in the native state. The importance of the cooperativity within this loop was also pointed out by another study of Hilser et al. [65].

The existence of these correlations suggests that close neighbor correlation of the rotations of the bonds may restrict the ensemble of conformations populated in the denatured state to a relatively smaller subset with native-like structure. A similar conclusion was reached in a recent study [2] where the propensities of each 20 amino acids to occur mono- and dipeptides mapped to discrete regions of the Ramachandran map were computed from proteins of known structure. The composites of local structure propensities provided evidence for the local encoding of long-range structure by emphasizing the existence of important local side-chain–backbone interactions, which may severely restrict the ensemble of conformations populated in the denatured state.

3.7. Structural-based analysis of the conformations by Gaussian Network Model (GNM)

3.7.1. Fluctuations in the fast modes of dynamics

The residues, which are important for the folding and stability of the protein, can be projected from the fast mode analysis of GNM as shown in recent studies [50,54]. The fluctuations of C^{α} atoms in several conformations extracted from the trajectories at different times of the simulation are calculated by GNM. The residues that exhibit the highest fluctuations in the fastest 10 modes (Eq. (7)) in all simulations, restrained and unrestrained, are shown in the ribbon diagram of CI2 in Fig. 7(a) and the fluctuations of the residues in the fastest 10 modes for several conformations are presented in Fig. 7(b). The peaks of the fluctuations in the fastest mode of dynamics, which are considered to be highly localized [66] are mainly located in the hydrophobic core; at some regions of α-helix, between Ile29-Ile37 of the hydrophobic small region in the middle of the chain and around the C terminus of β_2. Strikingly, the residues Ala16, Leu49, and Ile57 that are considered as the nucleation site for the folding of CI2 [21] exhibit the highest fluctuations in all simulations. These residues appear to maintain relatively high total number of contacts throughout the simulations, yet the distances between these residues increases significantly after TS, as shown in the MD simulations of Kazmirski et al. [21]. Overall, the residues obtained from the fast mode analysis are good candidates to be the key residues for folding of CI2, i.e. to be the extended nucleus of CI2. [67–69]. We would like to note that these residues (Ala16, Leu49, Ile7) exhibit lowest flexibility (i.e. minima near 0 values) in the slowest modes.

3.7.2. Fluctuations in the slowest mode of dynamics

The mean-square fluctuations of the residues along the most dominant and the cooperative mode (the first slowest mode) of the dynamics (Eqs. (5) and (7)) are calculated for the conformations with various Q values from the native to the unfolded states by GNM. The slowest mode shapes of several conformations chosen as representatives for given ranges of Q, which are plotted in Fig. 8(a)–(d), display how the mode shapes change as the structure travels from the native to the denatured. In the native state, the minimum fluctuating regions, i.e. hinge-points, correspond to the residues of the N terminus of the α-helix, residues around the middle of β_1 and most of the residues of the β_2, some of the latter being the most persistent. On the other hand, the inset in the Fig. 8(a) presents the mean-square fluctuations of C^{α} atoms obtained from X-ray crystallographic temperature factors in comparison with those obtained by GNM using all modes of the dynamics (Eq. (7)), which are in good agreement. The good prediction of the latter fluctuations also by MC simulations using the present model and potential of interactions employed in the present simulations was given in a previous study [37]. As can be observed, the

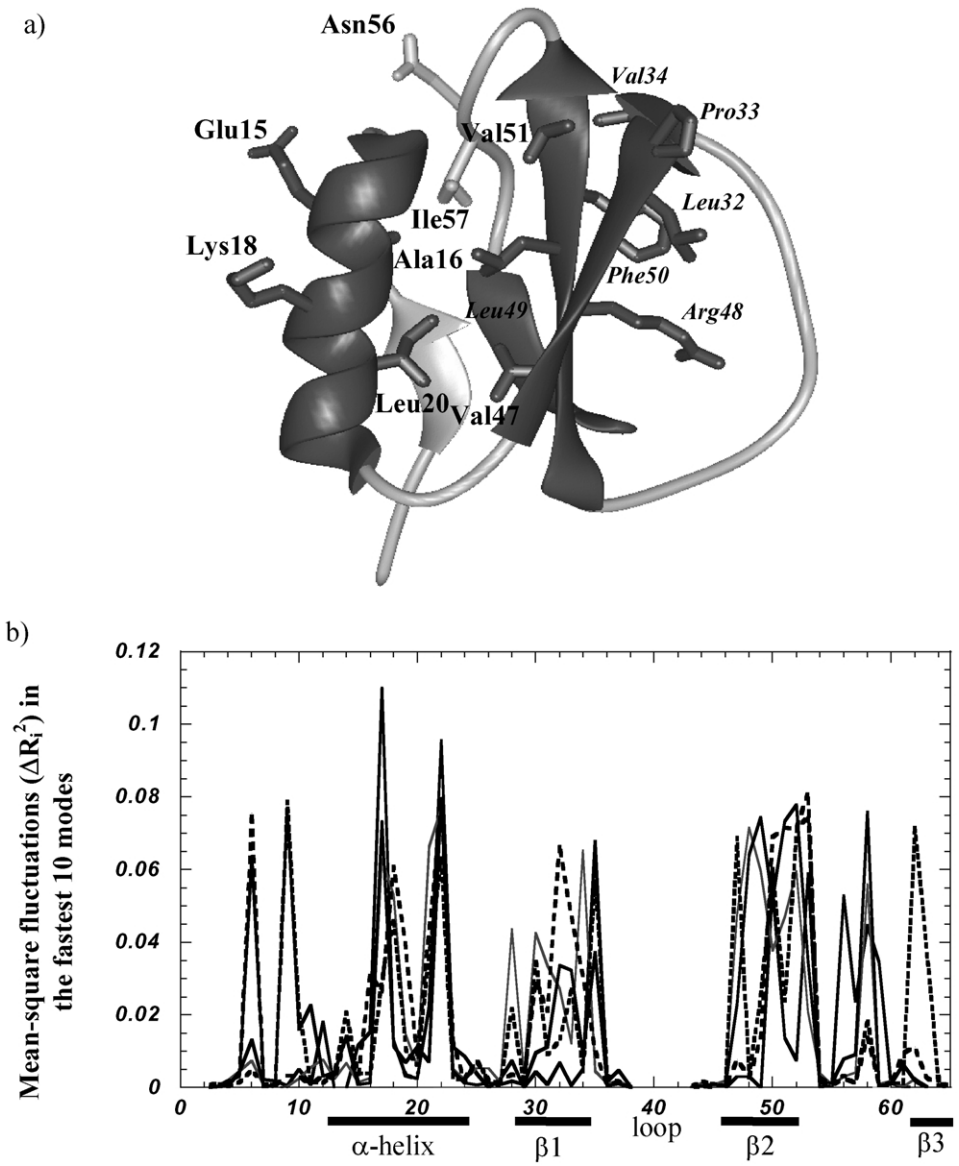

Fig. 7. The ribbon diagram showing the residues exhibiting the highest fluctuations in the fastest modes of dynamics (a), and the fluctuations of the residues in the fastest 10 modes (b), by GNM analysis of the several conformations visited during the unfolding.

fluctuations in the slowest dynamic mode are a good approximation to the behavior of the fluctuations obtained in the superimposition of all modes.

By analyzing the change in the most dominant mode shape of the structures as the fraction of the native contacts diminishes it can be seen that, the most persistent native-like hinge seems obviously to be the one around residues in the β_2; whereas the one at the N terminus of the α-helix is lost first, followed by the one in the middle of the β_1. The hinge at the middle of β_1 has been relatively persistent because of the continuing contacts between β_1 and β_2. Interestingly, the residues around the C terminus of the β_2 appear to be the most conserved residues of the structure, as shown above. The hinge around the residues of the β_2 separates the structure into two domains which show negatively correlated vibrational fluctuations, as discussed in the forthcoming section. The dashed and dotted curves (Fig. 8(d)) that were obtained by the average of all conformations within $0.2 < Q < 0.1$ show a hinge point near active site of CI2 (Met 40-Glu 41). Interestingly, in experimental studies of CI2, Niera et al. [70] separated the protein into two fragments by cutting at Met 40-Glu41 and observed that the two fragments reunite into CI2. It is notable that, if we further unfold the structure into more extended conformations, the hinge moves to the left of the β_2 towards to the loop, and then to the middle of the chain, where the chain behavior can simply be approximated by the Rouse chain model [71] in which only near neighbor correlations along the sequence are emphasized (see inset in Fig. 8(d)).

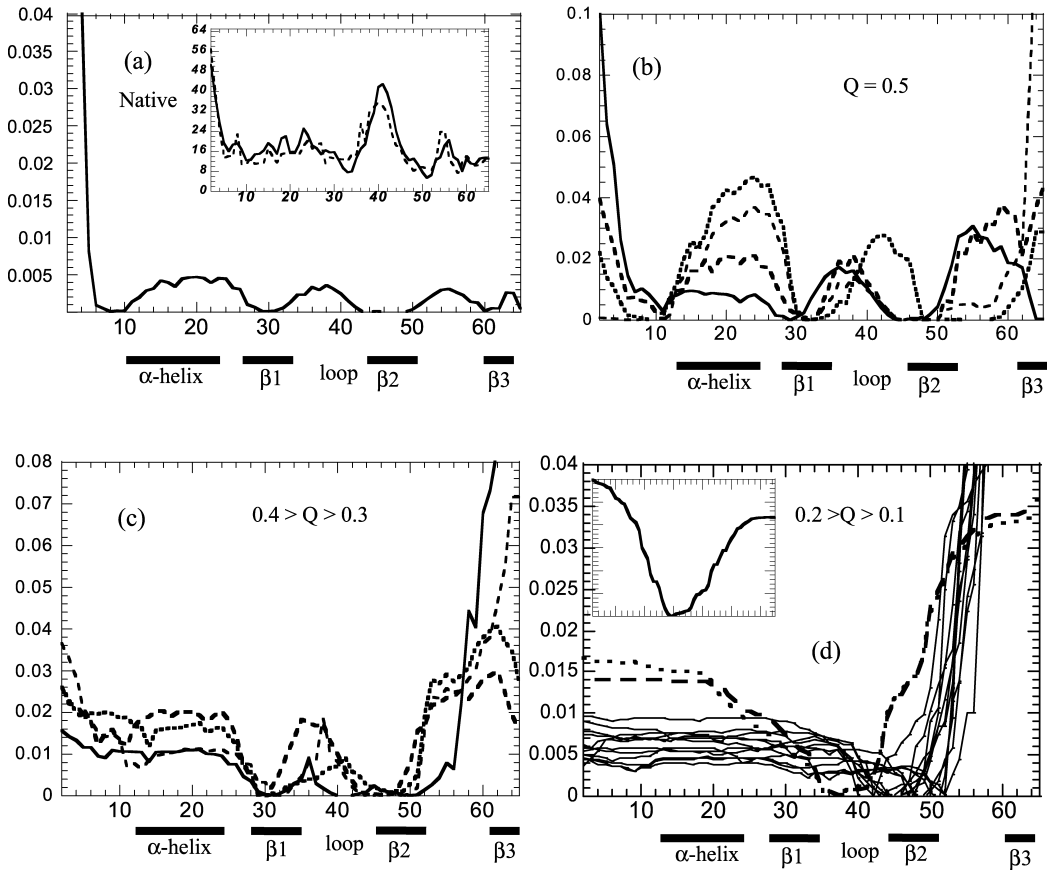

Fig. 8. (a)–(d) The mean-square fluctuations of each residue in the slowest mode of dynamics of the conformations with varying Q values, visited during the unfolding (with average S_g and RSMD values of (a) 11.04 (S_{g0}); (b) 12.4 (1.12 S_{g0}), 14.25; (c) 14.22 (1.28 S_{g0}), 17.44; and (d) 20.22 (1.83 S_{g0}), 23.5; respectively), by GNM analysis. S_{g0} is the native state radius of gyration. The inset in part (d) of the figure depicts the slowest mode shape of residue fluctuations in a highly extended conformation with S_g and RMSD values of 36.9 (3.3 S_{g0}) and 48.05, respectively. The mode shape approaches in this limit that of the Rouse chain model [72].

3.7.3. Correlations between the fluctuations of residues

Correlations between the fluctuations of residues are obtained by a structure-based analysis of the conformations from the unfolding trajectories by GNM. Fig. 9(a)–(d) depicts the correlation maps of the fluctuations of residues, including all modes of motion, for the conformations with various Q values. The two axes in the correlation map refer to residue indices, and the value on the map for a given pair of residues i and j is the correlation coefficient C_{ij} between the fluctuations of these residues, normalized with respect to their mean-square fluctuations, i.e.

$$C_{ij} = \frac{\langle \Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_j \rangle}{[\langle \Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_i \rangle \cdot \langle \Delta \mathbf{R}_j \cdot \Delta \mathbf{R}_j \rangle]^{1/2}} \tag{11}$$

The maps are symmetric, because $C_{ij} = C_{ji}$. Contours are obtained by connecting points of equal correlation. C_{ij} values vary in the range $-1 \leq C_{ij} \leq 1$. The lower limit refers to fully anticorrelated—same direction but opposite sense—fluctuations, and the upper limit to fully correlated—same direction and same sense—correlations. Uncorrelated, weakly correlated or perpendicular fluctuations yield C_{ij} values equal or close to zero.

The analysis of the correlations of the fluctuations in the conformations visited during the unfolding yields the following observations along with a set of representative maps chosen for the respective states. (Fig. 9(a)–(d)). S_g and RMSD of each conformation are given in the figure caption. Fig. 9(a) presents the correlation of the fluctuations in the native state, i.e. in the crystal structure at time 0. Along the diagonals, green/yellow/red blocks indicate the contiguous sequences of positively correlated residues. (See the color code on the right column).These blocks exactly coincide with the structural domains of the α-helix, reactive-site loop and β-strands. We also observe positive correlations between the strands $\beta_1-\beta_2$, $\beta_2-\beta_3$ and between N terminal and β_3 that agree with the contact maps. The survival of positive correlations between β_1 and β_2 is observed at even low Q values, but diminishes as the Q value decreases further. The residual α-helical structure and the small hydrophobic clusters of residues close along the sequence appear at even relatively low Q values. In general, the positive correlations of the residue fluctuations along the diagonal become more pronounced; whereas stronger negative correlations—which indicate the fluctuations in the opposite directions—between the residues of α-helix-loop ($\beta_1-\beta_2$), α-helix–β_3, and α-helix–β_2, are observed

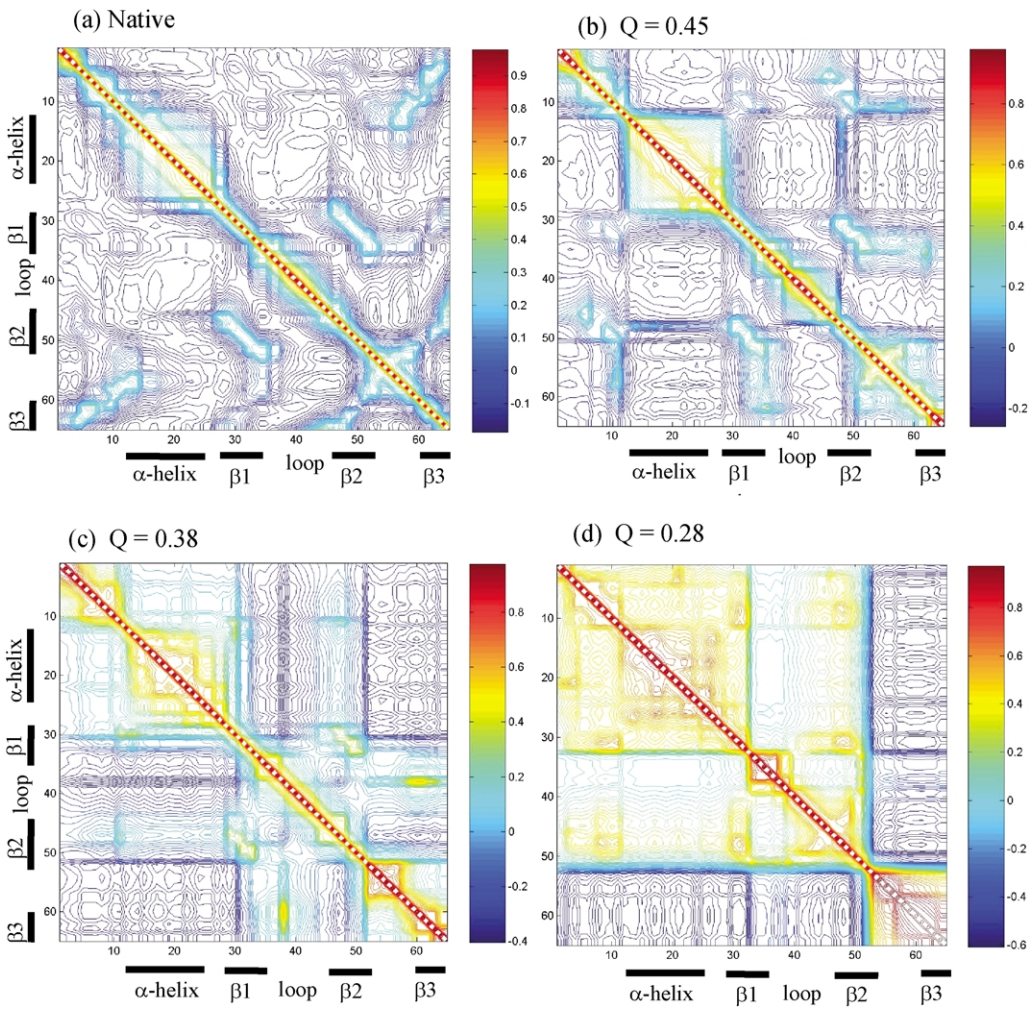

Fig. 9. (a)–(d) The correlations between the fluctuations of residues observed in some selected conformations with varyious Q values, visited during the unfolding (with S_g and RMSD values of (a) 11.04 (S_{g0}); (b) 12.1 (1.12 S_{g0}), 19.4; (c) 13.9 (1.26 S_{g0}), 16.0; (d) 17.12 (1.55 S_{g0}), 24.55; respectively), by GNM analysis.

towards lower Q values. Recent studies [72,73] emphasize the importance of cooperative interactions that reduce the population of intermediates during the unfolding transitions of two-state systems. The termini of the chain seem to undergo correlated fluctuations with non-native contacts as a result of the freely moving ends. At highly extended conformations, the structure appears to be separated into two domains the vibrational fluctuations of which were negatively correlated by a native-like hinge at residues around the C terminus of β_2. If the chain is further forced to attain more extended conformations where all the native and non-native contacts are lost, the Rouse chain model with only nearest-neighbor local interactions describes the collective motions of the two domains separated at the middle of the chain (inset in Fig. 9(d)). The native-like hinge at around β_2 indicates the importance of specific interactions dictated by the sequence enforcing the chain's behavior to deviate from the most general behavior, which is characteristic of the any flexible chain, to find native-like behavior and initiate folding.

Overall, the picture shows that as the chain unfolds, and relatively local interactions become more pronounced. Dokholayan et al. [74] analyzed pre- and post-transition structures of CI2 using a similar approach. They calculated the average graph connectivity of the structures based on their contact network topology. They also observed that post-transition structures have more long-range interactions which creates a more cooperative network.

The results from this structure-based analysis of the conformations and direct analysis of the unfolding trajectories suggest that highly correlated structural dynamics may be a universal intrinsic property of a polypeptide chain that has the ability to fold into a native protein, as recently discussed along with the results from QENS (quasielastic neutron-scattering) on α-lactalbumin, where both the residual secondary structure and tertiary-like long-range interactions are observed to fluctuate over several hundred picoseconds both in the native and in the denatured states [73]. Although, for two-state folding proteins, the folding kinetics can be predicted using simple, empirical,

structure-based rules rather than requiring the finest level of details of the structure [75], the amino acid sequence of proteins can stabilize not only the final state but also a small set of discrete partially folded conformations as discussed by Rumbley et al. [76].

4. Conclusions

An efficient and simple off-lattice MC simulation technique based on a low-resolution model with statistically driven potential of interaction has been used to carry out unfolding simulations of CI2 to search for its unfolding behavior following from the denatured state. Our results indicate a preferred pathway for the unfolding of CI2: it starts with the loss of native contacts between N terminus and β_3 and continues with β_2 and β_3. The denatured state of CI2 is a highly unfolded; the persistent β_1 and β_2 contacts are lost as the distance from the TS increases with diminishing Q. Some helical contacts and some non-native contacts around the termini persist even in highly unfolded conformations of the structure. The search of the cooperativity in the bond rotations revealed that certain regions of the structure; the α-helix, C terminus of the β_1-part of the reactive site loop and the residues around the C terminus of β_2, depict short-range cooperative bond rotational correlations along the sequence in even highly denatured conformations. The bonds around the C terminus of β_2 also display long-range cooperative rotations with other parts of the structure at intermediate values of Q. Notably, Φ-values that are obtained from simulations are in quite good agreement with those of experimental values, having a Pearson correlation coefficient of 0.68. This also serves to validate the extent of the reliability of the method.

In addition to direct analysis of the unfolding trajectories, the structure-based analysis of the conformations visited during unfolding allows us to search how the fluctuations of residues and their correlations change as the native contacts are lost during unfolding. The results show that the denatured state does not behave as a random coil, but instead seems to have an ability to exhibit negatively correlated fluctuations between the two major domains of the structure separated by the native-like hinge at around β_2, along with the partly folded α-helix and a small hydrophobic cluster in the middle of the chain. The residues in the latter regions also display high frequency fluctuations. The hinge-like motion observed in the slowest dynamics mode supports the existence of long-range correlations in the denatured state [3]. If the chains have highly extended conformations, this hinge-point moves to the middle of the chain, where the behavior can simply be defined as that of the Rouse model chain in which only nearest neighbor interactions exist. This suggests that, the sequence code of the chain, by means of local interactions, which is well defined by the short-range rotational cooperative motions of the virtual bonds along the chain, help the chain undergo a selective conformational search that culminates in finding its native like-topology. This consideration would automatically define initiation for the folding. The correlated motions of termini with nearest-neighbor residues may be the result of non-native contacts of the freely moving ends (as apparent with the contact analysis). The correlated fluctuations between two domains may induce favorable interactions between β_1 and β_2, which is extremely important for the folding of the structure, as discussed in a recent study [30].

Acknowledgements

This work was partially supported by BU Research Funds, grant numbers 00HA502D-00HA503 and by DPT Project 01K120280. This work has been also supported by the Turkish Academy of Sciences, in the framework of the Young Scientist Award Program (EA-TUBA-GEBIP/2001-1-1).

References

[1] Plaxco KW, Gross M. Nat Struct Biol 2001;8:659.
[2] Shortle D. Protein Sci 2002;11:18.
[3] Shortle D, Ackerman MS. Science 2001;293:487.
[4] Klein-Seetharaman J, Oikawa M, Grimshaw SB, Wirmer J, Duchardt E, Ueda T, Imoto T, Smith LJ, Dobson CM, Schwalbe H. Science 2002;295:1719.
[5] Anfinsen C. Science 1973;123:223.
[6] Fersht AR. Curr Opin Struct Biol 1995;5:79.
[7] Baldwin RL. J Biomol NMR 1995;5:103.
[8] Creighton T. Philos Trans R Soc London B Biol Sci 1995;348:5.
[9] Bryngelson JD, Onuchic JN, Socci ND, Wolynes PG. Proteins 1995; 21:167.
[10] Chan HS, Dill KA. Proteins 1998;30:2.
[11] Shakhnovich EI. Curr Opin Struct Biol 1997;7:29.
[12] Pande VJ, Grosberg AY, Tanaka T, Rokhsar DS. Curr Opin Struct Biol 1998;8:68.
[13] Thirumalai D, Klimov DK. Curr Opin Struct Biol 1999;9:197.
[14] Dinner AR, Sali A, Smith LJ, Dobson CM, Karplus M. Trends Biochem Sci 2000;25:331.
[15] Ozkan SB, Bahar I, Dill KA. Nat Struct Biol 2001;9:765.
[16] Hunenberg PH, Mark AE, Van Gunsteren WF. Proteins 1995;21: 196–213.
[17] Boczko EM, Brooks III CL. Science 1995;269:393.
[18] Ferrara P, Apostolakis J, Caflisch A. Proteins 2000;39:252.
[19] Gilgiun B, Guilbert C, Perahia D. Proteins 2000;41:58.
[20] Daggett V, Li A, Itzhaki LS, Otzen DE, Fersht AR. J Mol Biol 1996; 257:430.
[21] Kazmirski SL, Wong K, Freund SMV, Tan Y, Fersht AR, Daggett V. Proc Natl Acad Sci USA 2001;98:4349.
[22] Lazaridis T, Karplus M. Science 1997;278:1928.
[23] Zhou Y, Karplus M. Nature 1999;402:400.
[24] Dokholyan NV, Buldyrev SD, Stanley HE, Shakhnovich EI. J Mol Biol 2000;296:1183.
[25] Li L, Mirny LA, Shakhnovich EI. Nat Struct Biol 2000;7:336.
[26] Socci ND, Onuchic JN, Wolynes PG. Proteins 1998;32:136.
[27] Pande VJ, Rokhsar DSD. Proc Natl Acad Sci USA 1999;96:9062.
[28] Klimov DK, Thirumalai D. Proc Natl Acad Sci USA 2000;97:2544.
[29] Klimov DK, Thirumalai D. J Mol Biol 1998;282:471.

[30] Li L, Shakhnovich EI. Proc Natl Acad Sci USA 2001;23:13014.
[31] Matouscheck A, Kellis JT, Serrano L, Fersht A. Nature 1989;340:122.
[32] Matouscheck A, Serrano L, Fersht AR. J Mol Biol 1992;224:819.
[33] Clementi C, Nymeyer H, Onuchic JS. J Mol Biol 2000;298:938.
[34] Nymeyer H, Socci ND, Onuchic JS. Proc Natl Acad Sci USA 2000;97: 634.
[35] Takada S. Proc Natl Acad Sci USA 1999;96:11698.
[36] Haliloglu T, Bahar I. Proteins 1998;31:271.
[37] Kurt N, Haliloglu I T. Proteins 1999;37:454.
[38] Kurt N, Haliloglu T. J Biomol Struct Dyn 2001;18:713.
[39] McPhalen CA, James MN. Biochemistry 1987;26:261.
[40] Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The protein data bank. Nucl Acids Res 2000;28:235.
[41] Fersht AR. Structure and mechanism in protein science: a guide to enzyme catalysis and protein folding. New York: WH Freeman; 1999.
[42] Bahar I, Jernigan RL. Fold Des 1996;1:357.
[43] Flory PJ. Statistical mechanics of chain molecules. New York: Hanser Publishers; 1998.
[44] Bahar I, Jernigan RL. J Mol Biol 1997;266:195.
[45] Bahar I, Kaplan M, Jernigan RL. Proteins 1997;29:292.
[46] Bahar I, Atilgan AR, Demirel MC, Erman B. Phys Rev Lett 1998;80: 2733.
[47] Tirion MM. Phys Rev Lett 1996;77:1905.
[48] Amadei A, Linssen AB, Berendsen HJC. Proteins 1993;17:412.
[49] de Groot BL, Hayward S, van Aalten DMF, Amadei A, Berendsen HJC. Proteins 1998;31:116.
[50] Bahar I, Jernigan RL. J Mol Biol 1998;281:871.
[51] Bahar I, Jernigan RL. Biochemistry 1999;38:3478.
[52] Keskin O, Jernigan RL, Bahar I. Biophys J 2000;78:2093.
[53] Haliloglu T, Bahar I, Erman B. Phys Rev Lett 1997;79:3090.
[54] Bahar I, Wallquist A, Covell DG, Jernigan RL. Biochemistry 1998; 37:1067.
[55] Haliloglu T, Bahar I. Proteins 1999;37:654.
[56] Michelleti C, Lattanzi G, Maritan A. J Mol Biol 2002;321:909.
[57] Shen T, Canino LS, McCammon JA. Phy Rev Lett 2002;89: 068103–1068103.
[58] Micheletti C, Banavar JR, Maritan A. Phy Rev Lett 2001;87:088102.
[59] Day R, Benion JB, Ham S, Daggett V. J Mol Biol 2001;322:189.
[60] Rao F, Cafflisch A. J Chem Phys 2003;119:4035.
[61] Canet D, Lyon CE, Scheek RM, Robillard GT, Dobson CM, Hore PJ, van Nuland NAJ. J Mol Biol 2003;330:397.
[62] Clementi, Jenings PA, Onuchic JN. J Mol Biol 2001;311:879.
[63] Li A, Daggett V. J Mol Biol 1996;257:412.
[64] Itzhaki LS, Otzen DE, Fersht AR. J Mol Biol 1995;254:260.
[65] Hilser VJ, Dowdy D, Oas TG, Freire E. Proc Natl Acad Sci USA 1998;95:9903.
[66] Demirel MC, Atilgan AR, JernigaN RL, ErmanB I. Bahar Protein Sci 1998;7:2522.
[67] Fersht AR. Proc Natl Acad Sci USA 2000;97:1525.
[68] Ptitsyn OB, Ting KL. J Mol Biol 1999;291:671.
[69] Ting KL, Jernigan RL. J Mol Evol 2002;54:425.
[70] Niera JL, Davis B, Ladorner AG, Buckle AM, Gay GD, Fersht AR. Fold Des 1996;1:189.
[71] Rouse PE. J Chem Phys 1953;21:1272.
[72] Hilser VJ, Freire E. Proteins 1997;27:1671.
[73] Bu Z, Cook J, Callaway DJ. J Mol Biol 2001;312:865.
[74] Dokholayan NV, Li L, Ding F, Shakhnovich EI. Proc Natl Acad Sci USA 2002;99:8637.
[75] Plaxco KW, Simons KT, Ruczinski I, Baker D. Biochemistry 2000; 39:11177.
[76] Rumbley J, Hoang L, Mayne L, Englander SW. Proc Natl Acad Sci USA 2001;98:105.