PROTEINS: Structure, Function, and Genetics 37:454–464 (1999)

Conformational Dynamics of Chymotrypsin Inhibitor 2 by Coarse-Grained Simulations

Neşe Kurt and Türkan Haliloğlu*
Polymer Research Center and Chemical Engineering Department, Boğaziçi University, Bebek, Istanbul, Turkey

ABSTRACT **A coarse-grained dynamic Monte Carlo (MC) simulation method is used to investigate the conformational dynamics of chymotrypsin inhibitor 2 (CI2). Each residue is represented therein by two interaction sites, one at the α-carbon and the other on the amino acid side-chain. The energy and geometry parameters extracted from databank structures are used. The calculated rms fluctuations of α-carbon atoms are in good agreement with crystallographic temperature factors. The two regions of the protein that pack against each other to form the main hydrophobic core exhibit negatively correlated fluctuations. The conformational dynamics could efficiently be probed by the time-delayed orientational and conformational correlation functions of the virtual bonds: the active site loop, excluding the active site bond, the turn region, and the N-terminal of the α-helix are relatively more mobile regions of the structure. A correlation is observed between the hydrogen/deuterium (H/D) exchange behavior and the long-time orientational and conformational autocorrelation function values for CI2. A cooperativity in the rotations of the bonds near in sequence is observed at all time windows, whereas the cooperative rotations of the bonds far along the sequence appear at long time windows; these correlations contribute to the stability of the secondary structures and the tertiary structure, respectively. Proteins 1999;37:454–464.**

Key words: low resolution model; knowledge-based potentials; dynamic Monte Carlo; fluctuations; virtual bond rotations; orientational and conformational correlations

INTRODUCTION

The low-resolution models and methods to characterize protein conformational dynamics are of great interest to explore motions of the order of nanoseconds to milliseconds that cannot be spanned by conventional fully atomistic molecular dynamics (MD) simulations. Motions of these time-scales are generally collective conformational changes: a) the correlated fluctuations between sequentially distant but topologically coupled residues, b) the spatial reorganization and/or structural transition of secondary structure units, c) the cooperative changes in the tertiary contacts, and d) the larger-scale motions such as domain movements, all of which contribute to the eventual disruption of the folded structure. Conventional MD simulations with all atomic coordinates with full hydration are feasible only for the near-native fluctuations of small proteins at early stages of unfolding but at the cost of excessive computational times. Further, MD simulations of nanosecond calculations are still far from converged for pair correlations of low-frequency motions because of the problem of undersampling. On the other hand, simulations on a coarse-grained scale became efficient and physically sensible with the use of knowledge-based potentials extracted from the x-ray– and NMR-elucidated protein structures to study the collective motions. Original single-site-per-residue model and potentials[1,2] have been followed by more detailed potentials at which the distance[3,4] and direction[5–7] dependence of nonbonded interactions and the torsional preferences of backbone bonds were included.[8,9]

In general, the low-resolution models and parameters have been used for exploring the equilibrium properties of proteins, such as discriminating between correct and incorrect folds in threading experiments,[10–12] predicting the effects of amino acid substitutions on stability,[13] and recognizing peptide-binding sites on proteins.[14] On the other hand, the dynamics of proteins, using low-resolution models, have been attempted by a few studies only, after the pioneering work by Levitt and Warshel.[15] Most of the later studies have been in the form of on-lattice simulations.[16–23] Yet, some studies in recent years depict the utility of off-lattice simulations.[23–29]

In the present study, a recently developed off-lattice dynamic Monte Carlo (MC) simulation technique[30] using a coarse-grained model and energy parameters developed by Bahar et al.[4,9] has been used to explore the dynamics of a small monomeric protein, chymotrypsin inhibitor 2 (CI2). This protein has been addressed previously by both experimental[31,32] and simulation studies.[33] The twofold objective of the present study is to show the applicability of the present simulation technique by comparing some of the results with those of MD simulations and experiments and to further complement the findings accumulated by the latter studies with comparably longer simulation trajectories at much less computational time.

CI2, also known as barley serine proteinase inhibitor 2, is an eighty-three–residue protein from the albumin frac-

Grant sponsor: Boğaziçi University Research Foundation; Grant number 97A0502.
*Correspondence to: Türkan Haliloğlu, Polymer Research Center, Boğaziçi University, 80815 Bebek, Istanbul, Turkey. E-mail: turkan@prc.bme.boun.edu.tr
Received 23 November 1998; Accepted 18 June 1999

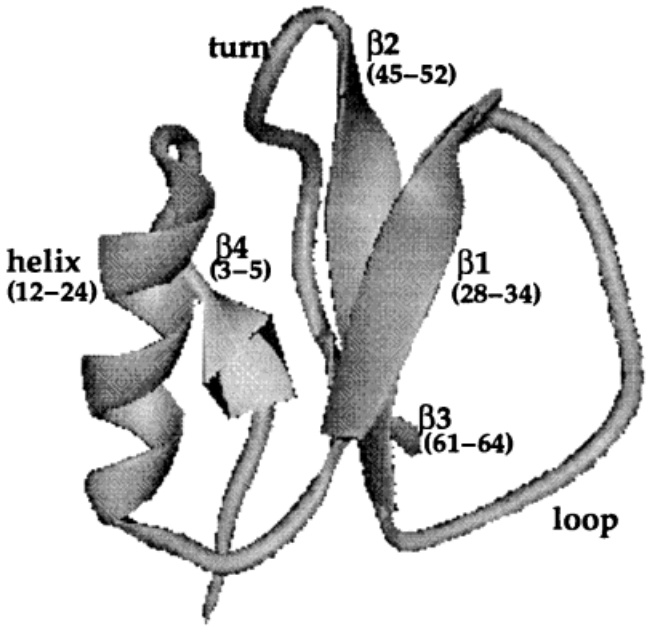

Fig. 1. Ribbon representation of the main-chain fold of CI2. The diagram is based on the crystal structure with PDB code 2ci2 determined to 2-Å resolution.[34] The residues were renumbered to begin with 1 instead of 20 for ease of comparison with previous studies on 2ci2. The residues forming the secondary structure elements are indicated in paranthesis according to this numbering scheme.

tion of barley seeds. The coordinates of the first nineteen residues are not resolved by either x-ray crystallography or 2D NMR. It is known that these residues do not contribute to the stability and activity of the protein. In this work, the atomic coordinates of CI2 determined by x-ray crystallography[34] to 2.0-Å resolution were obtained from the Brookhaven PDB, and the residues were renumbered beginning from one for the twentieth residue to exclude the disordered N-terminal region. The original numbering is also given in parenthesis for ease of comparison with previous studies on CI2.

The secondary structure of CI2 consists of a single α-helix with twelve residues 12–24 (31–43) and four β-strands: β1 residues 28–34 (47–53), β2 residues 45–52 (64–71), β3 residues 61–64 (80–83), and an extended strand β4 at the N-terminal [residues 3–5 (22–24)] running antiparallel to β3. There is a broad loop between strands 1 and 2 that contains the reactive site bond, between Met40 (59) and Glu41 (60). A ribbon representation of the structure is shown in Figure 1.

In this article we summarize the coarse-grained model with its energy functions and the simulation method in Model and Method. In Results and Discussion, the mean-square auto- and cross-fluctuations of α-carbon atoms and orientational and conformational correlation functions of the virtual bonds are presented and discussed with analogous observations from experiments. Conclusions follow.

MODEL AND METHOD

In the simplified model used, every residue is represented by two sites, its α-carbon atom and a sidechain interaction center S_i specific to the amino acid type.[7] The backbone of the protein is represented by the virtual bond model originally proposed by Flory and collaborators.[35] Accordingly, the backbone of a protein of n residues consists of n-1 virtual bonds connecting successive α-carbons. The conformation of the backbone is thus defined by 3n-6 variables: n-1 virtual bond lengths, l_i, connecting the α-carbons i-1 and i, n-2 virtual bond angles θ_i at the ith α-carbon, and n-3 dihedral angles ϕ_i describing the torsional rotation of the bond l_i. The side-chain of the residue is represented by a single interaction site, S_i, which is selected on the basis of specific properties of the amino acid type. S_i is found from the centroid of either all side-chain atoms or a few specific atoms, depending on the hydrophobicity or polarity of the amino acid. Details are in Bahar and Jernigan.[4,7]

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 chain sequence, and long-range interaction potentials $E_{LR}[\Phi]$, between non-bonded residues that are close in space.

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

Here $E_{LR}[\Phi]$[4] 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 sites i and j in conformation [Φ]. $W_{SS}(r_{ij})$ 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 was calculated as described by Bahar and collaborators[9] using the formulation

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

The first summation in the formula refers to the potential associated with the stretching of the virtual backbone bonds, approximated by a stiff harmonic potential with a force constant of 10 RT/Å^2. The second term is to account for distortion of bond angles, and the third is for the bond torsions, in which ϕ_1^- 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 the pairwise interdependence of the torsion and/or bond angle distortions.

The protein structure is allowed to move by differentially perturbing a randomly chosen backbone or side-chain interaction site. The perturbation causes a move given by the formula

$$\Delta x = k(2r - 1) \quad (4)$$

where r is a random number generated between 0 and 1 and k is a proportionality factor. This proportionality factor chosen controls the strength of perturbation and is chosen as 0.8 Å, which allows the protein to move only in the neighborhood of the original conformation. The move is accepted or rejected according to the Metropolis criterion. One MC step is composed of N = 64 moves; this may be viewed as the average time for all N residues of CI2 to have a chance to move. In the present simulations, *eight* independent runs of 4,800 MC steps each have been carried out. The lengths of MC trajectories were observed to lead to an rmsd of the final structure from crystal structure coordinates of an average of 4.0 Å. Each run requires not more than 2 hours CPU time on a Silicon Graphics R4400 Challenge Workstation and that is faster by about two orders of magnitude compared with the time required for a comparable MD trajectory. The average acceptance rate is 60%. The coordinates of backbone and side-chain interaction sites were recorded every eight MC steps. The subsequent analysis of the trajectories was made over 600 snapshots of each run, and the results were averaged over all runs.

RESULTS

Mean-Square Fluctuations in Atomic Coordinates

The mean-square fluctuations $\langle \Delta R_i^2 \rangle$ in the position vectors R_i of the α-carbon atoms around the crystal coordinates, averaged for eight independent simulations, are shown together with those calculated from crystallographic temperature factors in Figure 2. The results are normalized so that the area under the curve equals unity, to make possible a direct comparison. The simulation predicts the fluctuations at the minima and at the peaks in the reactive site loop [residues 34–45 (53–64)] and in the turn [residues 52–55 (71–74)] regions. Yet, the fluctuations of α-carbon atoms in the α-helix are found to have lower values than those measured. The same kind of discrepency was also observed in the MD simulations of Li and Daggett.[36] However, it is also notable in the latter study that the MD-derived average structure of the protein does not fall entirely within the envelope defined by the 20 NMR structures, especially in the loop region, which is successfully predicted with the present simulations. This discrepency was attributed[36] to the single MD simulation, which might not be adequate. The high mobility of the C-terminus is consistent with the high-amplitude fluctuations measured by an NMR study of Ludvigsen et al.,[37] in which the N- and C-termini and the loop region of the protein were reported as having the highest mean-square fluctuations.

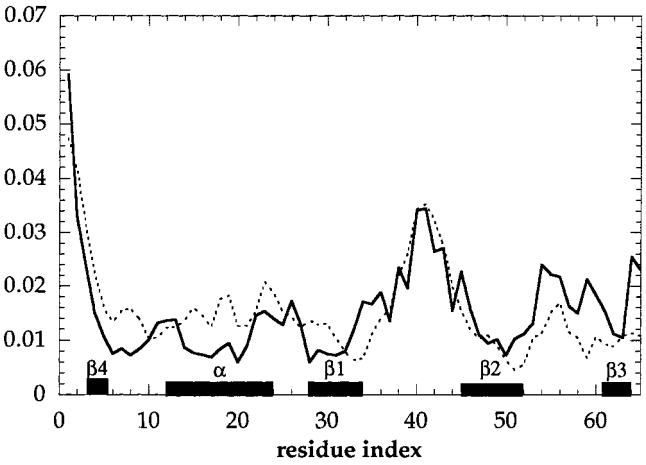

Fig. 2. The mean square fluctuations $\langle \Delta R_i^2 \rangle$ in α-carbon positions as a function of residue numbers for CI2. The solid and dashed curves depict the results from simulations and from experimental crystallographic temperature factors, respectively. The results are normalized.

Cross-Correlations Between Residue Motions

Equilibrium cross-correlations between residue motions were evaluated using the quantity $C_{ij} = \langle \Delta R_i \cdot \Delta R_j \rangle / (\langle \Delta R_i^2 \rangle \langle \Delta R_j^2 \rangle)^{1/2}$. Being normalized by the mean-square fluctuations of individual residues, C_{ij} gives an estimate of the orientational correlation between the motions undergone by sites i and j, eliminating the effect of the amplitude of motion. Figure 3 displays the residue pairs exhibiting negative correlations in the range $-1 \leq C_{ij} \leq -0.3$ (upper triangle) and positive correlations in the range $0.3 \leq C_{ij} \leq 1$ (lower triangle). The ordinate and abscissa represent the residue indices i and j. The regions of secondary structural elements are depicted on both axes.

The β-strands 1 and 2 are positively correlated with each other and with the reactive site loop, and negatively correlated with the α-helix. This suggests that these two strands and the loop form a group that undergoes negatively correlated fluctuations with the α-helix. In fact, Li and Daggett[36] divided the protein into two regions: the first region consists of the N-terminus and the α helix (residues 1–23), and the second region consists of the β-sheet and the loop (residues 23–64). These two regions pack against each other to form the hydrophobic core. They found that the packing of the two regions was destabilized and the core was weakened in the early stages of unfolding in their MD simulations. The negative correlation observed in the present simulations between these two groups also suggests that they make fluctuations in opposite directions, which would cause them to move apart in an unfolding process and weaken the hydrophobic core.

Furthermore, the residues in the C-terminus of the α-helix are positively correlated with each other. They are found to make cooperative fluctuations in the same direc-

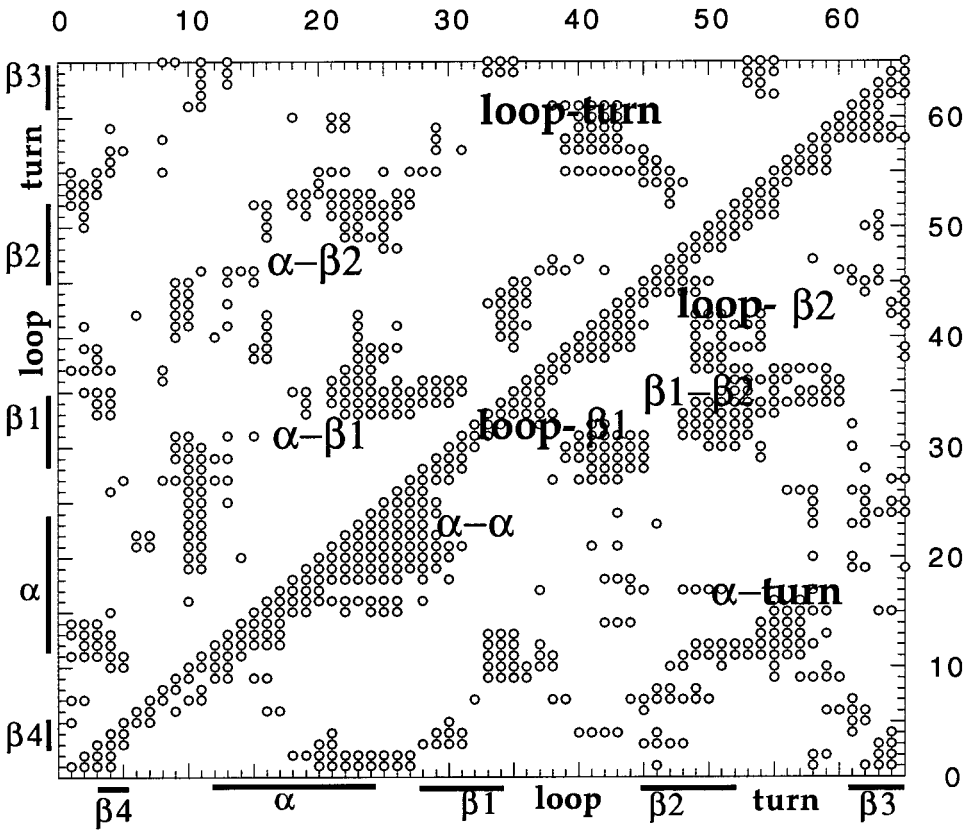

Fig. 3. Correlation map for the fluctuations of α-carbon atoms in CI2. The axes represent residue indices $1 \le i \le N$ and $1 \le j \le N$, where $N = 65$. The equilibrium cross-correlations are found from $C_{ij} = \langle \Delta R_i \cdot \Delta R_j \rangle / (\langle \Delta R_i^2 \rangle \langle \Delta R_j^2 \rangle)^{1/2}$. The diagonal and the lower triangular parts display the regions exhibiting positively correlated fluctuations, and the upper triangular portion refers to negatively correlated (or anti-correlated) pairs of residues.

tion, which means this terminus is more stable in this structural element. No correlation between the neighboring residues in the N-terminus of the α-helix is observed. However, the N-terminus displays positive correlations with the turn region. These two elements can make cooperative motions that will cause the top of the β-sheet to move away from the loop in any case of unfolding process, as observed in the MD simulation of Li and Daggett.[36] The negative correlation between the loop and turn regions also supports this idea.

Orientational and Conformational Correlations

The time-delayed correlations between the conformational states of residues are evaluated by analyzing the evolution of rotational states of backbone virtual bonds. The torsional autocorrelation function is defined as

$$G_i(\tau) = \langle \cos[\phi_i(t + \tau) - \phi_i(t)] \rangle. \quad (5)$$

Where $\phi_i(t + \tau)$ and $\phi_i(t)$ are the rotational angles of the ith bond at respective times $t + \tau$ and τ. The "ith bond" refers to the virtual bond between i^{th} and $i - 1^{th}$ α-C atoms. Figure 4**a** depicts $G_i(\tau)$ values at $\tau = 400$, 1,200, and 2,000 MC steps. The three curves are similar in character, although the losses of conformational correlations become more pronounced at longer times. Residues belonging to the same secondary structure are indicated by the same symbol for clarity.

The virtual bonds at the α-helix and β-sheet are observed to be highly autocorrelated, compared with those of the loop and turn regions. The N-termini of the α-helix and the β-strands are more mobile than their C-termini, and they all have a mobile residue in the middle that enhances the rotational mobility of the virtual bond at that point. Although the N-terminal residues of the α-helix are known to form the folding core,[38,39] the relationship between the folding and hydrogen exchange pathways is questionable.[40] Neira et al.[40] found that all the residues in the first two turns of the helix undergo local exchange, which indicates that they are rotationally mobile and exhibit local structural fluctuations. This agrees with the results displayed in Figure 4, which illustrate that the N-terminal of the α-helix is mobile and shows an unwinding tendency.

On the other hand, the virtual bonds of the loop and turn regions between the α-helix and the β-strands show a fast decay in rotational autocorrelations. Especially the N-terminal portion of the loop and the turn between strands 2 and 3 are remarkably mobile, in agreement with MD results.[36] It is interesting to note that the active site bond in the loop is less mobile than the rest of the loop, and that the loop is separated into two portions by the active site, the N-terminal portion being more mobile than the C-terminal portion. This accords with the ^{15}N relaxation data measurements of Shaw et al.[41] These authors pointed out that the N-terminal portion of the loop becomes

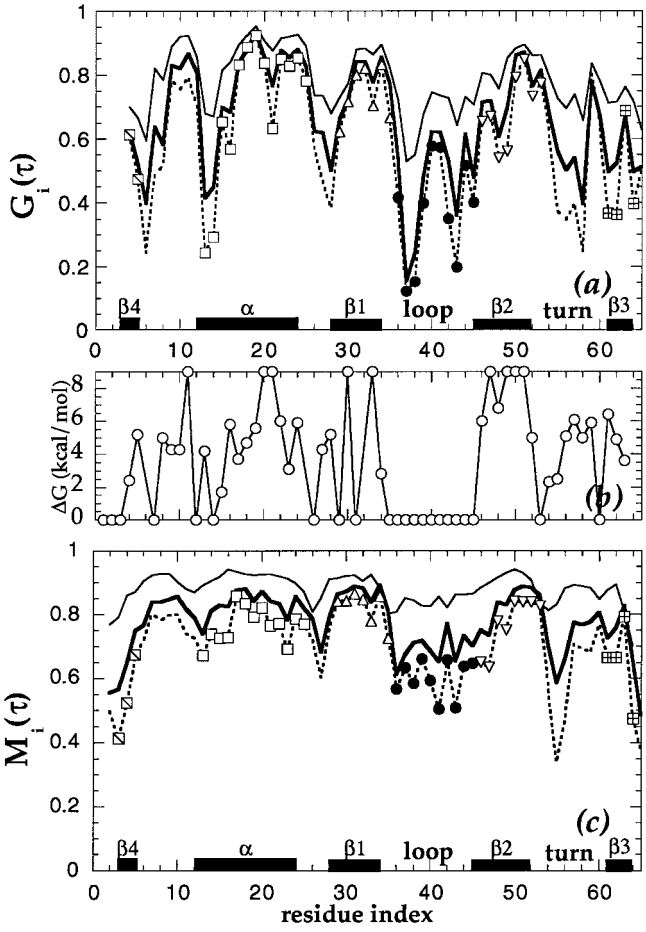

Fig. 4. (**a**) Time-delayed conformational autocorrelations, $\langle\cos[\phi_i(t+\tau) - \phi_i(t)]\rangle$, for the rotations of all virtual bonds in CI2 at τ = 400 (solid thin), 1,200 (solid thick), and 2,000 (dashed) MC steps. The bonds belonging to secondary structural elements are indicated by different symbols for clarity. (**b**) H/D exchange data from Neira et al.[40] as a function of residue index. (**c**) Time-delayed orientational autocorrelations, $\langle l_i(t) \cdot l_i(t+\tau)\rangle$, for the virtual bonds in CI2. The curves are for the same MC steps as in part (a).

disordered whereas the C-terminal region remains relatively ordered upon cleavage of the active site bond.

The two regions of the protein that pack against each other to form the main hydrophobic core are connected to each other by the residues between the α-helix and the β-sheet 1. These residues have high conformational mobility, which may be the result of negatively correlated motions of these two regions (Fig. 3). Daggett et al.[36] discussed how this turn is extended in the transition state to accommodate the movement of the α-helix away from the core.

The experimental free energy changes, ΔG, for hydrogen/deuterium (H/D) exchange in CI2[40] are displayed in Figure 4**b.** A correlation is observed between the H/D exchange data and long-time conformational and orientational autocorrelation function values, the results of the latter functions being presented below. Thus, the local flexibility induced by the torsional motions emerges as a property that affects the rate of H/D exchange. As far as the long-time behavior is concerned, because the time in MC simulations is tentative, to have an analogous sense of the real-time scale, the relaxation time of orientational behavior (correlation time of a second order correlation function, the second order Legendre polynomial) of a hypothetical bond that is appended perpendicular to the vertex of the two successive virtual bonds was calculated. The latter hypothetical bond was considered for different parts of the chain, and an average value of 100 MC steps was obtained for the relaxation time. Thus, long-time refers to 20 times the time necessary for a local relaxation—it should be pointed out here that a virtual bond comprises three real bonds in the structure—and to a trajectory of 48-fold for a statistically reliable analysis. On the other hand, the average correlation time of an N-H bond for the most constrained regions of CI2 was calculated as 160 ps in an MD study by Li and Dagget.[36] Consequently, it can be concluded that 100 MC steps, as being the characteristic time for a local relaxation that involves a segment of six real bonds (two successive virtual bonds), apparently represent a time scale greater than at least 160 ps.

The time-delayed orientational autocorrelation function of the virtual bond is expressed as

$$M_i(\tau) = \langle l_i(t) \cdot l_i(t+\tau)\rangle. \quad (6)$$

Where $l_i(t)$ is the unit vector along bond i at time t. The results for all virtual bonds for time intervals of τ = 400, 1,200, and 2,000 MC steps are depicted in Figure 4**c.** The results are qualitatively similar to those of torsional autocorrelations. This is expected, because the torsional motions affect the orientation of the virtual bonds. However, the decay of the orientational correlation functions is slower than that of the torsional correlation functions. In particular, the N-termini of the α-helix and β-strands 1 and 2 do not lose their bond orientations as fast, although they were found to undergo rapid torsional motions. This suggests the existence of cooperative torsional angle rotations that localize the motion. This property has also been observed for polymers in dense media.[42]

The most mobile part of the protein according to the losses of rotational correlations is the loop region, but it seems likely that the bonds in this region undergo some cooperative rotations because they preserve to some extent their original orientations (compare the ordinate values for bonds in the loop region in parts a and c of Figure 4), as will be elaborated in Cross-Correlations Between Bond Rotations, and the splitting of N- and C-terminal portions cannot be observed.

On the other hand, the turn between β-strands 2 and 3 is a mobile region of the protein that loses both its torsional and orientational correlations rapidly. This type I turn observed in the crystal structure was found to be not well defined in the solution structures by the NMR studies of Ludvigsen et al.[37] Their solution structure calculations showed that this region has some minor violations of both

the angular and the distance boundaries. Upon these observations, they suggested that this region may contain two or more conformations in the solution structure. These findings support the results that show that this region has a high mobility and would lose its orientational correlation even faster than the loop region. This mobile turn, the residues of which exhibit highly correlated fluctuations with those of the loop and α-helix, may have an important role in the overall cooperative motions of the protein, as will be emphasized in the cross-correlations of the rotations as well.

The cross-correlations between the bond rotations, which will be analyzed next, give some insight into the cooperativity in the rotations of the backbone bonds.

Cross-Correlations Between Bond Rotations

The time-delayed cross-correlations, $C_{ij}(\tau)$, between the rotational angles of the virtual bonds are calculated by using the expression

$$C_{ij}(\tau) = \langle \Delta\phi_i(\tau)\Delta\phi_j(\tau)\rangle / (\langle \Delta\phi_i(\tau)^2\rangle^{1/2}\langle \Delta\phi_j(\tau)^2\rangle^{1/2}) \qquad (7)$$

where $\Delta\phi_i(\tau)$ and $\Delta\phi_j(\tau)$ are the changes in the rotational angles of the virtual bonds i and j within the time τ.

The pairs of bond rotations having cross-correlations greater than 0.25 or less than −0.25 are depicted in Figure 5. The filled and open symbols refer to negative and positive cross-correlations, respectively. The strongly correlated pairs shown in **a** for smaller τ values (400 or 1,200 MC steps) are close to the diagonal, which means that they are neighboring virtual bonds. Strong correlations between the rotations of the virtual bonds that are separated far along the sequence can be observed at longer τ values (2,000 MC steps). This implies that coupled rotations at short time scales contribute to the stability on a local scale along the backbone. However, tertiary structure is supported by correlated rotational motions at longer time scales, as will be elaborated below.

Table I lists the pairs that exhibit correlated rotations with $C_{ij} > 0.20$ and $C_{ij} < -0.20$ at τ = 400 MC steps. The large number of correlated pairs explains why the orientational bond correlation functions of virtual bonds (Fig. 4a) have higher values compared with the correlations of torsional angles (Fig. 4b) at this time interval. The same pairs are observed to be correlated also at longer time intervals.

Figure 6**a** displays the time history of the torsional angles of the negatively correlated pair 13-14. These two bonds rotate in opposite directions to preserve the secondary structure by localizing the motion. This is also illustrated in Figure 4 by the significant difference between the torsional and orientational functions at all τ values of the N-terminus of the α-helix. For further illustrative purposes, the time history of rotational angles for the virtual bonds of 40, 42, and 44 in the binding loop are presented in Figure 6**b.** Rotations of bond 42 are highly negatively correlated with those of bonds 40 and 44. The negatively correlated rotations of these bonds localize the motion and

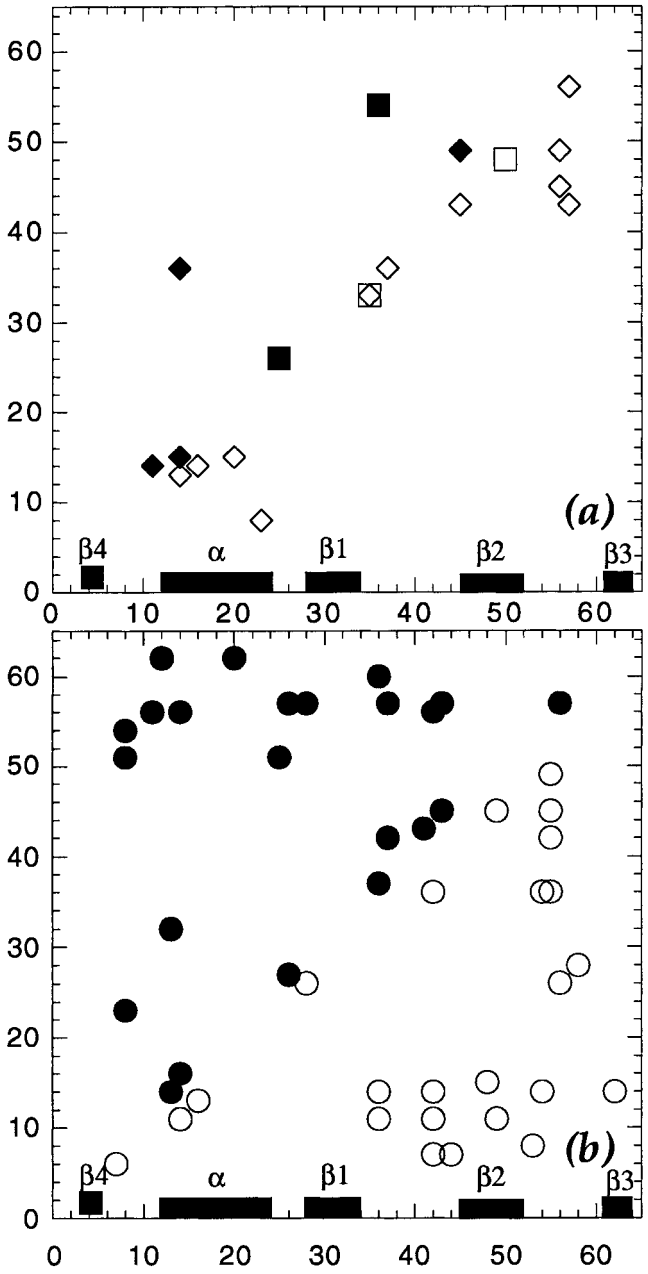

Fig. 5. Correlation map for the virtual bond rotations in CI2. Strongly correlated pairs at (**a**) τ = 400 (squares) or 1,200 (diamonds) MC steps. (**b**) τ = 2,000 (circles) MC steps. The open symbols are for positively correlated pairs ($C_{ij}(\tau) > 0.25$), and the closed symbols depict the negatively correlated ($C_{ij}(\tau) < -0.25$) virtual bond pairs.

cause the orientational autocorrelation functions to decay slower than the torsional autocorrelation functions of the rest of the bonds of the loop region (Fig. 4). The other negatively correlated bond pairs in this region are virtual bonds 35–37 and 43–45.

To illustrate the localization of the motion by the cooperative nature of the bond rotations for pair 13-14, a schematic diagram of the protein structure in the neighbor-

TABLE I: Pairs of Bonds Exhibiting Strongly Correlated Rotations in CI2, ($C_{ij} > 0.20$ and $C_{ij} < -0.20$) at τ = 400 MC steps

Negatively correlated pairs	Positively correlated pairs
13-14 (α)	6–7
29-31 (β1)	25–26
31-33 (β1)	28–29(β1)
33-35 (β1)	8–53
35-37 (loop)	
40-42 (loop)	
42-44 (loop)	
43-45 (loop)	
48-50 (β2)	

hood of this pair is drawn in Figure 6**c.** In this simplified diagram, two snapshots at t = 2,448 and t = 2,592 MC steps, presented by vertical dashed lines in Figure 6a, are superimposed. The torsional angle values for bonds 13-14 are indicated on the Figure for both snapshots. Despite the large amplitude rotations of these two virtual bonds, the localized motion does not lead to changes in the rest of the structure.

On the other hand, the long-time cross-correlations appearing between virtual bond pairs that are separated by at least two and usually by more than several virtual bonds along the sequence are the result of surviving collective rotational motions of several bonds. A 3D ribbon

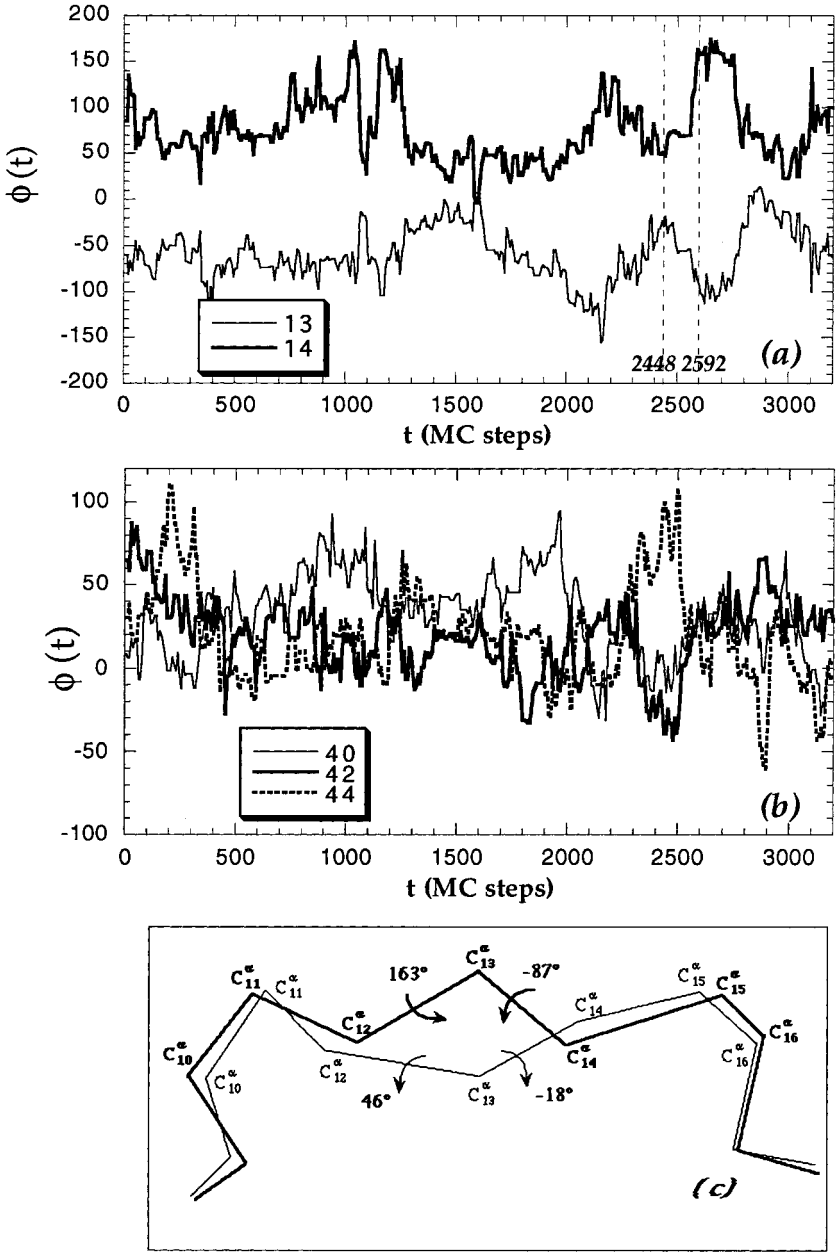

Fig. 6. (**a**) Time evolution of bond rotations for the virtual bond pair 13-14. The strong anticorrelation observed indicates the cooperative nature of bond rotations at the N-terminal of the α-helix in CI2. (**b**) Time evolution of bond rotations for the virtual bonds 40, 42, and 44. Rotations of bond 42 are negatively correlated with those of bonds 40 and 44. (**c**) Schematic representation of a segment that includes the virtual bond pair 13-14. Snapshots of the structure at t = 2,448 MC steps (thin line) and t = 2,592 MC steps (thick line), are superimposed. The torsional angle values for this bond pair at these two snapshots are indicated on the Figure.

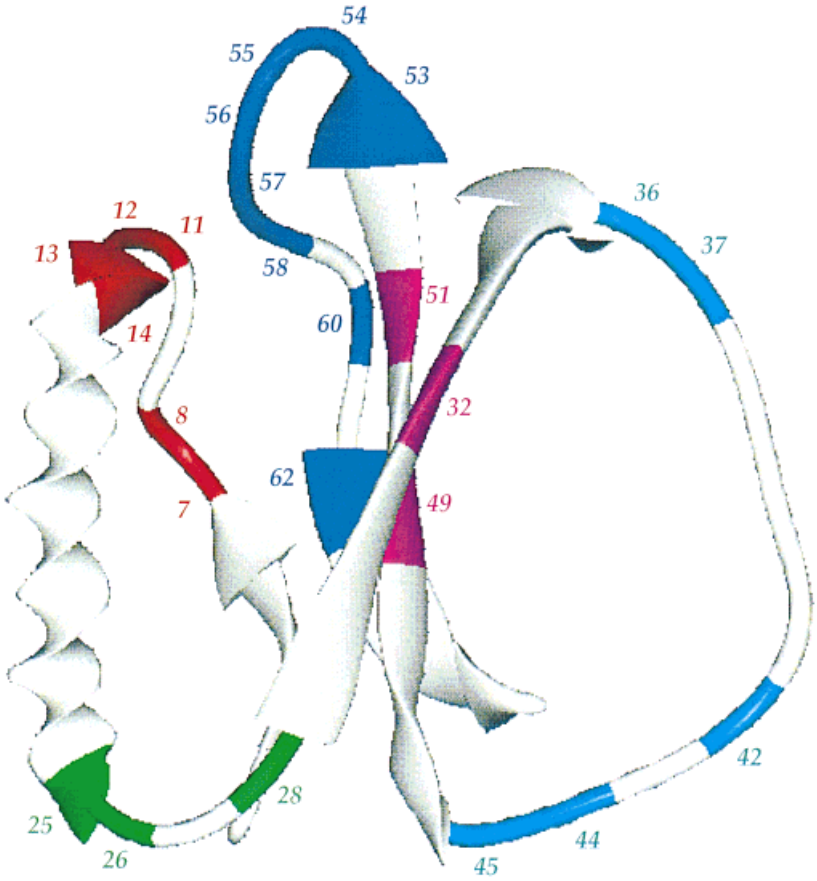

Fig. 7. A ribbon diagram of CI2 in which the bonds exhibiting highly coupled rotations at long time windows (see Table II) are labeled. The regions grouped in the text are colored as follows: I, red; II, green; III, cyan; IV, blue. The bonds that are not included in any of these groups are colored magenta.

TABLE II. Pairs of Bonds Exhibiting Strongly Correlated Rotations in CI2, $C_{ij} > 0.25$ (+) and $C_{ij} < -0.25$ (−), at τ = 2000 MC Steps

I-III	I-IV	II-IV	III-IV	Others
7-42 (+)	8-53 (+)	26-56 (+)	36-54 (+)	8-23 (−)
7-44 (+)	8-54 (−)	26-57 (+)	36-55 (+)	8-51 (−)
11-36 (+)	11-56 (−)	28-57 (−)	36-60 (−)	11-14 (+)
11-42 (+)	12-62 (−)	28-58 (+)	42-55 (+)	11-49 (+)
14-36 (+)	14-56 (−)		42-56 (−)	13-16 (+)
14-42 (+)	14-62 (+)		45-55 (+)	13-32 (−)
	14-54 (+)			25-51 (−)
				36-42 (+)
				37-42 (−)
				49-55 (+)

diagram on which most of the residues involved in correlated rotational motions are labeled is presented in Figure 7 to facilitate the understanding of their interactions. The cross-correlated rotational motions of the bonds revealed, in general, that the turns and the reactive site loop, in general, exhibit correlated motions among themselves. The N- and C-terminals of α-helix and N- or C-terminus of some β-strands are together with the adjacent turns in those correlated motions. Moreover, some bonds from α-helix, β_1, and β_2 are involved in the correlations as well.

To ease the discussion on the cross-correlations which appeared, the bonds exhibiting correlated rotations are grouped in regions as: (I) the turn between β_4 and α-helix and N-terminus of α-helix, (II) the turn between α-helix and β_1 and C-terminus of α-helix, (III) the loop (reactive site) between β_1 and β_2, and (IV) the turn between β_2 and β_3, C-terminus of β_2 and N-terminus of β_3. The residues that participated in the correlation motions belonging to the same regions are depicted with the same color. Table II depicts the correlated bond pairs ($C_{ij} > 0.25$ and $C_{ij} < -0.25$, Fig. 5**b**) that lead to the intercorrelated motions of the regions I–III, I–IV, II–IV, and III–IV at long times. Table 2 also depicts the correlated pairs involving the sites that are not defined in the latter regions. The sense of the correlations in the Table are depicted as (+) and (−) for the respective positive and negative rotational correlations. Region IV seems to be the only one in interaction with all three of the other regions. Region II is in interaction with IV only. Bond 49 in β_2 depicts relatively strong correlations with bonds 11(I), 55(IV), and 45(III), the latter being the strongest correlation observed in the present analysis. In addition to the interregional correlations defined above, correlated rotational motions are observed for the following pairs: 25 (C-terminus of α-helix)-51(β_2), which strengthens the interaction of II with the rest of the structure; 8(I)-23(α-helix); 8(I)-51((β_2); and

13(I)-32(β_1), which provides an interaction to I to some extent with α-helix, β_1 and β_2. On the other hand, correlation of bond pairs of 11–14 and 13–16, which might be viewed as short-range as well, appears strongly at long times in the N-terminus of α-helix. Further, in region III, which is the largest loop of CI2, in addition to short-range (nearest neighbor or next nearest neighbor pairs) correlations at short- and long-time windows, very strong intraloop correlated motion by bond pairs 36–42 and 37–42 appears at long times along with interregional correlations.

As an overall view from the analysis, it is notable that the rotational correlation of the bonds that are separated far along the sequence at long times do not necessarily have to be close in space; i.e., it is possible to observe a pair relatively far from each other but still in interaction via an another bond that is close in space and rotationally correlated to both bonds in the pair. For example, 55 is correlated with 36 and 42, 36 being much closer in space than 42. However, a strong correlation also exists between 36 and 42; consequently, a correlation is observed between 55 and 42. Additional examples are the interactions among 45, 49, and 55; 14, 36, and 42; 11, 36, and 42; and so forth.

To depict the cooperative rotations of some bond pairs, time histories of rotations for the bond pairs 36-42, 45-49, 8-51, and 14-36 are given in Figure 8**a–d.** Except for 8-51, which depicts negatively correlated rotations, the other pairs show positively correlated rotations at long time windows. A better understanding of the sense of the cooperativity in the rotations can be gained from the Figures by remembering that the long-time behavior corresponds to the change of the torsional angle of each bond in the pair in 250 snapshots. Therefore, the trend of the curves in that time scale should be followed.

The appearance of some correlations between the different regions of the structure is expected at long time windows because this type of pair-correlation is incorporating the collective motions of several residues in the vicinity and is consequently low-frequency motion in nature. Inasmuch as long-time behaviors are concerned, enough sampling of the occurrence of correlations is important for reliable analysis, as pointed out in a recent study[43] on the problem of undersampling in nanosecond MD simulation studies. The averaging over eight independent runs in the present simulations ensure the covering of the whole conformational space in addition to a relatively long simulation trajectory.

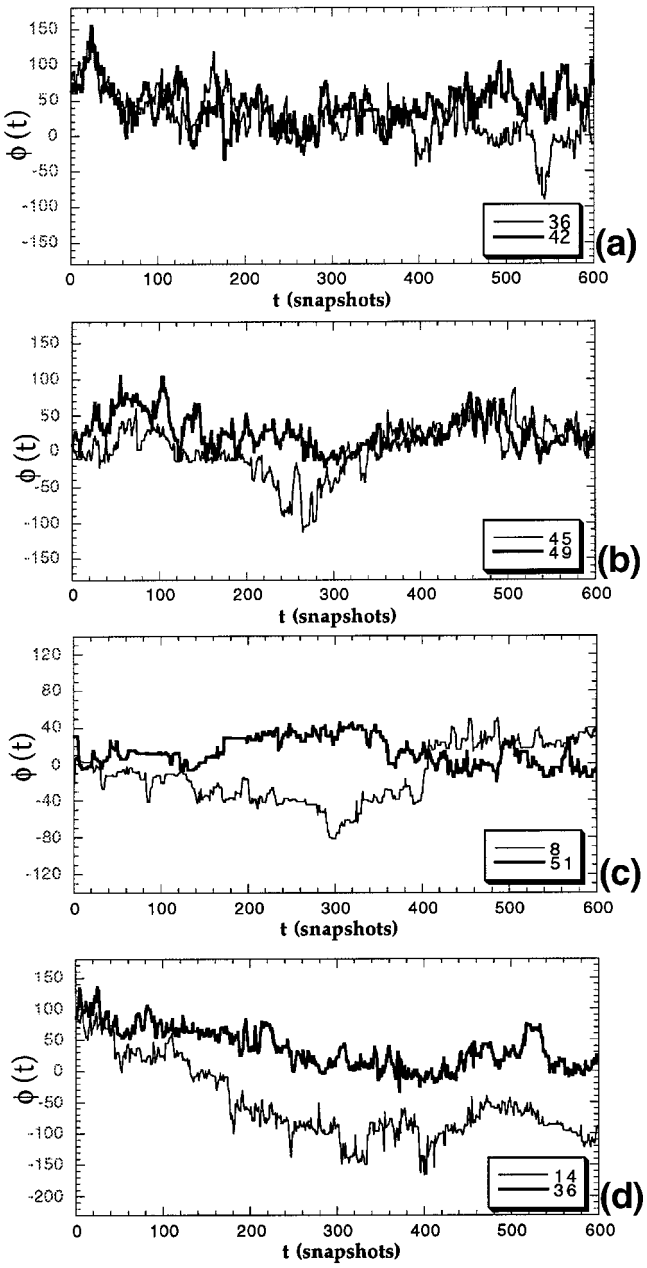

Fig. 8. Time history of rotations for some bond pairs that exhibit highly correlated torsions at long time windows. (**a**) 36-42, (**b**) 45-49, (**c**) 8-51, and (**d**) 14-36. Except for 8-51, which depicts negatively correlated rotations, the other pairs show positively correlated rotations.

CONCLUSIONS

We draw the following conclusions from the off-lattice MC simulations of inhibitor CI2.

A good agreement is obtained between the mean-square fluctuations in the position vectors of the α-carbon atoms and the experimental crystallographic temperature factors. The minima regions and peaks at the active site loop and the turn are satisfactorily predicted. The simulation predicts lower fluctuation values for the α-carbon atoms in the α-helix compared with those measured by experiments. This is in accordance with the MD simulation results.

The equilibrium cross-correlations between the fluctuations of the residue pairs permit identification of the units that are correlated. The two regions of the protein that pack against each other to form the main hydrophobic core exhibit negatively correlated fluctuations. This will cause

the hydrophobic core to be weakened in an unfolding process.

The conformational dynamics could be probed efficiently by the time-delayed conformational and orientational autocorrelation functions. In particular, the bonds of the active site loop, turn region, and N-terminal of the α-helix depict high-amplitude rotational motions. The active site bond is less mobile in this sense compared with the rest of the loop. The relaxation of the rotational motions of the bonds in the N-terminal portion of the loop is relatively faster than the relaxation of those in the C-terminal portion. In general, the conformational correlation functions decay faster than the orientational correlation functions because of the cooperativity between the rotations of the bonds.

A correlation exists between the H/D exchange data and the long-time conformational and orientational autocorrelation function values for C12. It may be concluded that the flexibility associated with the torsional motions affects the rate of H/D exchange.

Time-delayed cross-correlations between the bond rotations reveal the pairs of bonds that exhibit correlated rotational motions. The cooperativity in the rotations of bonds that are near neighbors along the sequence is observed at all time windows, which ensures the localization of the motion along the backbone and thus the stability of the secondary structures, whereas cooperative rotations of bonds that are far along the sequence, but close enough in space, are observed at longer time windows, which ensures the stability of the tertiary structure.

ACKNOWLEDGMENTS

We thank Professor Ivet Bahar for her motivating discussions.

REFERENCES

1. Tanaka S, Scheraga HI. Medium- and long-range interaction parameters between amino acids for predicting three-dimensional structures of proteins. Macromolecules 1976;9:945–950.
2. Miyazawa S, Jernigan RL. Estimation of effective inter-residue contact energies from protein crystal structures: quasichemical approximation. Macromolecules 1985;18:534–552.
3. Sippl MJ. Knowledge-based potentials for proteins. Curr Opin Struct Biol 1995;5:229–235.
4. Bahar I, Jernigan RL. Inter-residue potentials in globular proteins and the dominance of highly specific hydrophilic interactions at close separation. J Mol Biol 1997;266:195–214.
5. Singh J, Thornton JM. SIRIUS: an automated method of analysis of preferred packing arrangements between protein groups. J Mol Biol 1990;211:595–615.
6. Vriend G, Sander C. Quality control of protein models: directional atomic contact analysis. J Appl Crystallogr 1993;26:47–60.
7. Bahar I, Jernigan RL. Coordination geometry of non-bonded residues in globular proteins. Folding Design 1996;1:357–370.
8. DeWitte RS, Shakhnovich EI. Pseudodihedrals: simplified protein backbone representation with knowledge-based energy. Protein Sci 1994;2:1570–1581.
9. Bahar I, Kaplan M, Jernigan RL. Short-range conformational energies, secondary structure propensities, and recognition of correct sequence-structure matches. Proteins 1997;29:292–308.
10. Hendlich M, Lackner P, Weitckus S, et al. Identification of native protein folds amongst a large number of incorrect models: the calculation of low-energy conformations from potentials of mean force. J Mol Biol 1990;216:167–180.
11. Kocher JPA, Rooman MJ, Wodak S. Factors influencing the ability of knowledge-based potentials to identify native sequence-structure matches. J Mol Biol 1994;235:1598–1613.
12. Wang Y, Zhang H, Li W, Scott RA. Discriminating compact nonnative structures from the native structure of globular proteins. Proc Natl Acad Sci USA 1995;92:709–713.
13. Miyazawa S, Jernigan RL. Protein stability for single substitution mutants and the extent of local compactness in the denatured state. Protein Eng 1994;7:1209–1220.
14. Altuvia Y, Schueler O, Margalit H. Ranking potential binding peptides to MHC molecules by a computational threading approach. J Mol Biol 1995;249:244–250.
15. Levitt M, Warshel A. Computer simulation of protein folding. Nature 1975;253:694–698.
16. Ueda Y, Taketomi H, Go N. Studies on protein folding, unfolding and fluctuations by computer simulations. II. A three-dimensional lattice model for lyzosyme. Biopolymers 1978;17:1531–1548.
17. Skolnick J, Kolinski A. Dynamic Monte Carlo simulation of globular protein folding/unfolding pathway. I. Six member, Greek key β barrel proteins. J Mol Biol 1990;212:787–817.
18. Covell DG. Lattice model simulation of polypeptide chain folding. J Mol Biol 1994;235:1032–1043.
19. Kolinski A, Skolnick J. Monte Carlo simulation of protein folding. I. Lattice model and interaction scheme. Proteins 1994;18:338–352.
20. Sali A, Shakhnovich E, Karplus M. Kinetics of protein folding: a lattice model study of the requirements for folding to the native state. J Mol Biol 1994;235:1614–1636.
21. Vieth M, Kolinski A, Brooks ICL, Skolnick J. Prediction of the folding pathways and structure of GCN4 leucine zipper. J Mol Biol 1994;237:361–367.
22. Elofsson A, LeGrand SM, Eisenberg D. Local moves: an efficient algorithm for simulation of protein folding. Proteins 1995;23:73–82.
23. Karplus M, Sali A. Theoretical studies of protein folding and unfolding. Curr Opin Struct Biol 1995;5:58–73.
24. Sun S. Reduced representation model of protein structure prediction: statistical potential and genetic algorithms. Protein Sci 1993;2:762–785.
25. Park BH, Levitt M. The complexity of accuracy of discrete state models of protein structure. J Mol Biol 1995;249:493–507.
26. Sun S, Thomas PD, Dill KA. A simple protein folding algorithm using a binary code and secondary structure constraints. Protein Eng 1995;8:769–778.
27. Park BH, Levitt M. Energy functions that discriminate X-ray and near-native folds from well-constrained decoys. J Mol Biol 1996; 258:367–392.
28. Haliloğlu T. Characterization of internal motions of *Escherichia coli* ribonuclease H by Monte Carlo simulation. Proteins 1999;34: 533–539.
29. Haliloğlu T. Coarse-grained simulations of conformational dynamics of proteins. Theor Computat Polymer Sci 1999;9:255–260.
30. Haliloğlu T, Bahar I. Coarse-grained simulations of conformational dynamics of proteins: application to apomyoglobin. Proteins 1998;31:271–281.
31. Ladurner AG, Itzhaki LS, de Prat Gay G, Fersht AR. Complementation of peptide fragments of the single domain protein chymotrypsin inhibitor 2. J Mol Biol 1997;273:317–329.
32. Jackson SE, Moracci M, elMasry N, Johnson CM, Fersht AR. Effect of cavity-creating mutations in the hydrophobic core of chymotrypsin inhibitor 2. Biochemistry 1993;32:11259–11269.
33. Daggett V, Li A, Itzhaki LS, Otzen DE, Fersht AR. Structure of the transition state for folding of a protein derived from experiment and simulation. J Mol Biol 1996;257:430–440.
34. McPhalen CA, James MNG. Crystal and molecular structure of the serine proteinase inhibitor CI-2 from barley seeds. Biochemistry 1987;26:261–269.
35. Flory PJ. Statistical mechanics of chain molecules. New York: Hanser Publishers; 1998.
36. Li A, Daggett V. Investigation of the solution structure of CI2 using MD: comparison to x-ray crystallographic and NMR data. Protein Eng 1995;8:1117–1128.
37. Ludvigsen S, Shen H, Kjaer M, Madsen JC, Poulsen FM. Refinement of the three-dimensional solution structure of barley serine

proteinase inhibitor 2 and comparison with the structures in crystals. J Mol Biol 1991;222:621–635.
38. Otzen DE. The structure of the transition state for the folding/unfolding of the barley chymotrypsin inhibitor-2 and its implications for mechanisms of protein folding. Proc Natl Acad Sci USA 1994;91:10422–10425.
39. Itzhaki LS, Otzen DE, Fersht AR. The structure of the transition state for folding of CI2 analyzed by protein engineering methods: evidence for a nucleation-condensation mechanism for protein folding. J Mol Biol 1995;254:260–288.
40. Neira JL, Itzhaki LS, Otzen DE, Davis B, Fersht AR. Hydrogen exchange in CI2 probed by mutagenesis. J Mol Biol 1997;270:99–110.
41. Shaw GL, Davis B, Keeler J, Fersht AR. Backbone dynamics of CI2: effect of breaking the active site bond and its implications for the mechanism of inhibition of serine proteases. Biochemistry 1995;34:2225–2233.
42. Haliloğlu T, Bahar I, Erman B, Kim EG, Mattice WL. A dynamic rotational isomeric state approach for extension of the time scale of the local dynamics observed in fully atomistic molecular dynamics simulations: application to polybutadiene. J Chem Phys 1996;104:4828–4834.
43. Clarage JB, Romo T, Andrews BK, Pettitt BM, Philips GA. Sampling problem in molecular dynamics simulations of macromolecules. Proc Natl Acad Sci USA 1995;92:3288–3292.