Biochemical and Biophysical Research Communications 373 (2008) 64–68

Contents lists available at ScienceDirect

Biochemical and Biophysical Research Communications

journal homepage: www.elsevier.com/locate/ybbrc

Key residue-dominated protein folding dynamics

Xin-Qiu Yao [a,b], Zhen-Su She [a,b,c,*]

[a] State Key Laboratory for Turbulence and Complex Systems and Department of Biomedical Engineering, College of Engineering, Peking University, No. 5 Yihe Yuan Street, Haidan, Beijing 100871, China
[b] Center for Theoretical Biology, Peking University, Beijing 100871, China
[c] Department of Mathematics, University of California, Los Angeles, Los Angeles, CA 90095, USA

ARTICLE INFO

Article history:
Received 25 May 2008
Available online 10 June 2008

Keywords:
Key residue
Protein folding
Side-chain relaxation
Order parameter
Trp-cage

ABSTRACT

A "key-residue" hypothesis that a few residues' characteristics contain the essential dynamics of the whole protein is proposed for the study of side-chain relaxation near native states. Molecular dynamics simulation is performed on the folding of Trp-cage, and four key residues are discovered and shown to be highly sensitive to the change of state of the protein away from the native state. Order parameters that characterize the geometrical properties of key residues are shown to form valuable phase plane on which one distinguishes different reaction pathways. Furthermore, one of the key residues, Trp6, is observed to display two reconfiguration processes, in which one is induced by an unconstrained torsion of the side-chain of Trp6, with a rate faster by almost an order of magnitude than the other one described by Kussell's model. The faster process seems to occur more frequently in our simulation and thus represent a significant mechanism in folding dynamics.

Despite numerous efforts, the mechanism of protein folding is still not completely understood [1]. Most theoretical studies of protein folding have been restricted in coarse-grained protein models and in the mechanism of backbone folding. For example, a lattice protein model is used by Dill [2] to study the thermodynamic properties of protein, such as the phase-transition temperature and the dependency of stability on chain length. Shakhnovich et al. [3,4] have studied the nucleation-condensation mechanism of protein folding using both lattice and off-lattice models. Using a lattice model, Li et al. [5] have found the correlation between the symmetry of protein structure and the high designability of sequences. A comprehensive review of theoretical models for protein folding can be found in [1]. In principle, the models mentioned above are unable to deal with dynamical motions around native or near-native state, in which the atomic details of side-chains play a dominant role. At present, a satisfying theory accounting for both long-time behavior and sufficient atomic details of protein motions is still missing.

The side-chain dynamics and relaxation is recognized as a cornerstone of protein folding [6–9]. Actually, the excluded-volume effect, for which the exact organization of atoms is of particularly importance, is one of the key determinants of protein's tertiary structure [10]. Using Monte Carlo simulation, Kussell et al. have found that there are a large number of side-chain conformations compatible with a given native backbone, with 20% of χ angles in a non-native state for an average repacked conformation [7]. This result indicates that the relaxation process of side-chain is significant.

For side-chain relaxation, a completely different idea of coarse-graining may be derived by focusing on dynamical roles of heterogeneous residues: some residues are sensitive to the evolution of state of the protein and so are more relevant to the identification of proper reaction pathways, while the others are relatively insensitive. For the former category, which may be called "key residues", a detailed description of atomic dynamics is required to obtain an accurate folding picture; for the other "non-key residues", a simplification in both structure and interaction may be introduced. Present work based on all-atom molecular dynamics (MD) simulation has discovered four key residues for a small peptide called "Trp-cage" [11], using a simple classification rule. Based on key residues, the folding kinetics of Trp-cage is studied and two kinds of reconfiguration processes with significantly different reaction rates are discovered.

Materials and methods

The Trp-cage system is prepared from the NMR structure (PDB [12] entry 1l2y). A 3-ns unfolding ($T = 380$ K) simulation is performed at first, after which the protein reaches a compact denatured state: the radius of gyration for all heavy atoms (Rg) and

* Corresponding author. Address: State Key Laboratory for Turbulence and Complex Systems and Department of Biomedical Engineering, College of Engineering, Peking University, No. 5 Yihe Yuan Street, Haidan, Beijing 100871, China.
E-mail address: she@pku.edu.cn (Z.-S. She).

0006-291X/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.bbrc.2008.05.179

the root mean square deviation (RMSD) for all C_α atoms (denoted by C_α-RMSD) compared to the NMR structure are 0.79 and 0.38 nm, respectively. An important characteristic of the denatured state is that the six hydrophobic residues (Tyr3, Trp6, Leu7, Pro12, Pro18, and Pro19) that form the hydrophobic core in the NMR structure are incorrectly packed. Using the denatured state as initial conformation, the protein is then solvated in a box of SPC (simple point charge) water molecules, which contains 10,372 atoms in total. Before the production of data, an energy minimization process is carried out, followed by 1-ns equilibration for water molecules. Forty simulations of the folding process are performed in parallel, with the same initial conformations but with randomly selected initial velocities. The GROMACS package and OPLS/AA force field [13] are employed in conjunction, and the covalent bonds associated with hydrogen atoms are constrained using LINCS algorithm [14], allowing an integrating step-size of 2fs. Both Van de Waals' and electrostatic interactions are treated using cut-off method ($r = 1.2$ nm). All simulations are performed under the NPT (282 K, 1 atm) condition. The trajectory is saved every 10 ps, and after 100-ns integrations 10,000 conformations is collected for analysis in each simulation. To study the properties of the native state of the protein, a 40-ns simulation with the initial conformation being the NMR structure is also independently performed, denoted by native-MD.

Results and discussion

Detection of folding and mis-folding events

The native state of Trp-cage is characterized by the following four quantities: (i) C_α-RMSD; (ii) the radius of gyration of the hydrophobic core, denoted by HC-R_g; (iii) a stability score of backbone hydrogen bonds (HB) within the N-terminal α-helix, denoted by $HB_{\alpha\text{-helix}}$; and (iv) the side-chain dihedral angle χ_2 of Trp6, by Trp6-χ_2. The stability score of HB is defined by $\frac{n_{HB}}{N} \times 100\%$ and is calculated by running a sliding window (length = 500 ps) over the temporal trajectory, where N is the total number of conformations within the window and n_{HB} is the number of conformations in which the HB exists. Only the second backbone HB (i.e., between Tyr3 and Leu7) of the N-terminal α-helix is considered in the calculation because it is found (in the native-MD) to be relatively stable (average $HB_{\alpha\text{-helix}} > 50\%$) compared to other HB. Empirical statistics of all above quantities are derived from the native-MD, which are used as criteria to identify folding events: C_α-RMSD < 0.32 nm, HC-R_g < 0.48 nm, $HB_{\alpha\text{-helix}} > 53.5\%$ and $52^\circ <$ Trp6-$\chi_2 < 105^\circ$.

Multiple folding events are discovered from simulations using above criteria, and the fold conformations show close topology to the NMR structure (Fig. 1A). An interesting result is the observation of a "mis-fold" state, which has very similar characteristics to the true native state except for an incorrect placement of the side-chain of Trp6 (Trp6-χ_2 is about 180° difference between fold and mis-fold states, as shown in Fig. 1B). Note that previous study [15] has also found a similar "pseudo-native" state for Trp-cage; it seems that this protein may not be as stable as previously recognized by experiment [11].

In total, we have detected seven folding events and three mis-folding events, and the collections of corresponding trajectories are used to identify key residues, and to study the folding/mis-folding kinetics as well.

Identification of key residues

The present working hypothesis is that the high-dimensional folding process of Trp-cage can be understood by inspecting only the state of some key residues. We assumed that the key residues must be stable with little fluctuation in the native state while very unstable in non-native states, so that their temporal fluctuation patterns can clearly differentiate among the native and non-native states.

The relative position between C_α and an "end" atom is used to represent the orientation of the side-chain. The end atom is defined for all amino acids (excluding Gly) as following: Ala-C_β, Arg-$N\eta_2$, Asn-$N_{\delta 2}$, Asp-$O_{\delta 2}$, Cys-$S\gamma$, Gln-$N_{\varepsilon 2}$, Glu-$O_{\varepsilon 2}$, His-$N_{\varepsilon 2}$, Ile-C_δ, Leu-$C_{\delta 2}$, Lys-N_ζ, Met-$C\varepsilon$, Phe-C_ζ, Pro-$C\gamma$, Ser-$O\gamma$, Thr-$C\gamma_2$, Trp-$C\eta_2$, Tyr-$O\eta$, and Val-$C\gamma_2$. A deviation (D) from the NMR structure is defined for a residue in a given structure using the two representative atoms,

$$D = \sqrt{|\delta r|^2} = \sqrt{\frac{1}{2}\left(|r_{C_\alpha} - r^0_{C_\alpha}|^2 + |r_{end} - r^0_{end}|^2\right)} \quad (1)$$

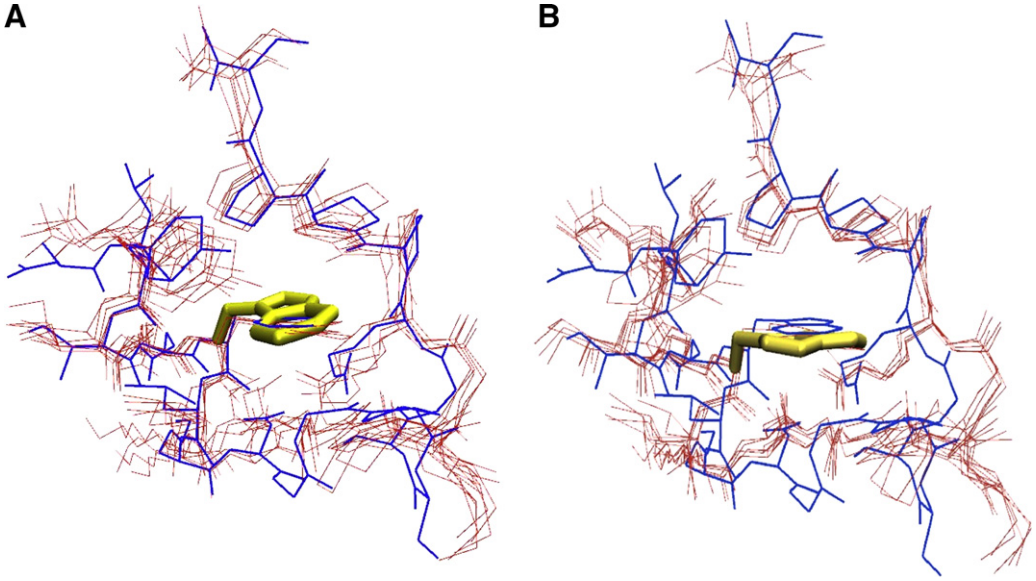

Fig. 1. View of five representative fold (A) and mis-fold (B) snapshots (red) superimposed on the NMR structure of Trp-cage (blue). The snapshots are taken from trajectories that reach fold and mis-fold state, respectively. The side-chain of Trp6 (yellow) is represented as thick bonds while other residues are represented as lines. Only heavy atoms are shown for clarity. (For interpretation of color mentioned in this figure the reader is referred to the web version of the article.)

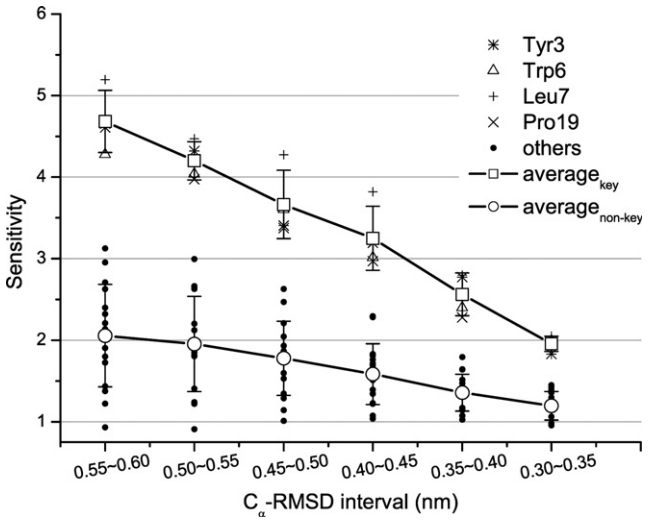

Fig. 2. Sensitivity of each residue to the change of state of the protein. The definition of sensitivity can be found in the text. The non-native state is further split into sub-ensembles according to the C_α-RMSD of each structure compared to the NMR structure, and for each sub-ensemble the sensitivities of all the 20 residues are calculated. Four residues, Tyr3 (star), Trp6 (triangle), Leu7 (plus) and Pro19 (cross) have clearly higher sensitivities than other residues (dot) for all sub-ensembles, and so the former are recognized as "key residues" and the latter as "non-key residues". The average sensitivity (with error bars) of each category is also calculated for each sub-ensemble.

where r is the atomic position after superimposition onto the NMR structure using all C_α atoms and r^0 is the position in the NMR structure. D is a measure of the "stableness" of each residue.

The "sensitivity" score of each residue with respect to the change of state of the protein is defined by

$$S_i = \frac{< D_i >_{\text{non-active}}}{< D_i >_{\text{native}}} \quad (2)$$

where i is the residue index and the average is taken on the ensemble of non-native and native conformations, respectively. The non-native state is further split into sub-ensembles according to the C_α-RMSD of each conformation compared to the NMR structure; this allows to reveal systematic variation of sensitivities defined above with respect to conformations of variable distances from the native state.

Fig. 2 shows the comparison of sensitivities of the 20 residues as a function of C_α-RMSD distances. There is a clear separation of symbols: the top four residues have persistently higher sensitivity score than other residues; the former consists of Tyr3, Trp6, Leu7, and Pro19, and are now recognized as key residues of Trp-cage. Fig. 2 also shows that the average sensitivity of key residues is larger than 2.0 for almost all C_α-RMSD sub-ensembles. On the contrary, the average sensitivity of "non-key" residues is generally below 2.0. Note that all four key residues are constituents of the hydrophobic core of the protein; this can be understood by the fact that Trp-cage is intentionally designed as a hydrophobic-force-driven folding protein [11].

Characterization of folding and mis-folding kinetics

The four key residues discovered above form a sub-system which is now characterized by a pair of order parameters: (i) the average distance between C_α atoms of key residues, denoted by L_{key}, and (ii) Trp6-χ_2. In this consideration, we take into account both translational and rotational aspects. The free energy landscape is calculated in the phase plane of L_{key} and Trp6-χ_2, as shown in Fig. 3A. On the landscape, the fold (labeled by "d") and mis-fold ("d'") states are clearly separated, locating at the positions (L_{key}, Trp6-χ_2) = (0.65 nm, 80°) and (0.65 nm, −110°), respectively. Both states have an energy level approximately $-8k_BT$, where k_B is the Boltzmann's constant and T is the temperature, indicating that the two states have equivalent stabilities.

The states explored by ensemble-averaged folding and mis-folding trajectories, represented by red and blue dots in Fig. 3A respectively, reveal quite different pathways for the two reactions. Specifically, the folding and mis-folding processes pass through different intermediate states, which are denoted by "b" and "b'", respectively. Both intermediate states have an energy level of $-6k_BT$, indicating that they are meta-stable compared to the fold or mis-fold state. The snapshots of structure taken along the reaction paths (Fig. 3C) indicate that in both "b" and "b'", the four key residues form a non-native hydrophobic cluster.

For both folding and mis-folding processes, there is an on-path transition state, denoted by "Ts" and "Ts'" respectively, as shown in Fig. 3A. The temporal direction of the reaction, as indicated by the arrows shown in Fig. 3A for each path, shows that there is a significant increase of L_{key} after passing the transition state, and then the protein reaches the "c" (or "c'") state. The increase of L_{key} corresponds to a particular dynamical event, which involves the loosening of local backbone of key residues. This event can be understood by inspecting the structure shown in Fig. 3C: the temporarily formed non-native hydrophobic cluster limits the state space that Trp6-χ_2 can access due to the excluded-volume effect; only the destruction of the cluster, and then the increase of L_{key}, can make room for Trp6-χ_2 to rotate into the native (or mis-fold) state.

Fig. 3A indicates that the loosening of backbone may undergo quite different dynamics at different state of Trp6-χ_2. For 120° < Trp6-χ_2 < 210° (−150°) and −10° < Trp6-χ_2 < 10°, a little increase of L_{key} will cost a relatively large amount of energy, and so the local backbone is stable; these regions are named as "constrained" as shown in Fig. 3A. On the other hand, for −150° < Trp6-χ_2 < −130° and 100° < Trp6-χ_2 < 120° the contour of free energy is stretched along L_{key}, and so the backbone can easily loosen without costing too much energy; these regions are named as "unconstrained". The classification of "constrained" and "unconstrained" regions in the free energy landscape indicates a significant correlation between the dynamics of Trp6-χ_2 and L_{key}, and this is investigated in the following section.

For comparison, we have also calculated the free energy landscape and the folding/mis-folding pathways using two traditional order parameters: the radius of gyration of all heavy atoms (denoted by R_g) and the fraction of native contacts (FNC). In our definition of contact, the pair of residues must be three-residue apart in the sequence and the smallest distance between any pair of heavy atoms from the two residues is less than 0.6 nm. Compared to Fig. 3A, both the reaction paths and the equilibrium states are difficult to distinguish between folding and mis-folding in Fig. 3B, indicating that the two traditional order parameters are unsuitable to represent the dynamics. The dynamics involves subtle placement of side-chains of key residues as dominant effects during motions near the native state, and hence are critical to the construction of proper order parameters.

Side-chain-induced reconfiguration

The apparent difference between the free energy distribution in "constrained" and "unconstrained" regions indicates that the state of Trp6-χ_2 has a significant effect on the statistical dynamics of L_{key}. A study of the correlation between the dynamics of Trp6-χ_2 and L_{key} is carried out by calculating following two time periods. The first is a time that Trp6-χ_2 stays in the "constrained" region from the beginning of simulation, as defined by $\Delta t_c = t_c - t_h$, where t_h is the first passage time (FPT) that L_{key} < 0.65 nm and t_c is the FPT

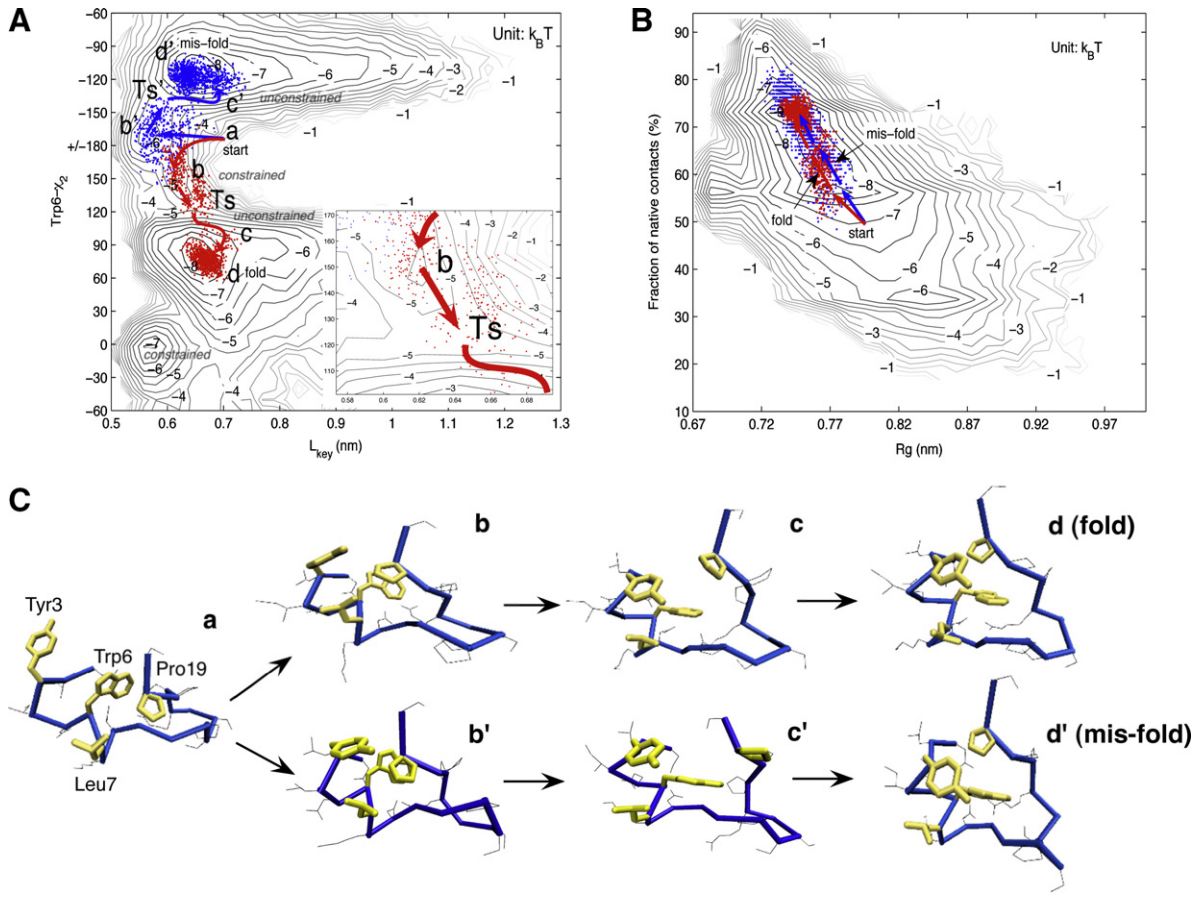

Fig. 3. Free energy landscapes and folding/mis-folding pathway. (A) Free energy landscape calculated using key-residue-based order parameters. The regions denoted by "constrained" and "unconstrained" are explained in the text. The folding and mis-folding pathways are represented by red and blue dots, respectively, and the temporal direction of each reaction is indicated by arrows. The key states that each reaction passes by are indicated by specific labels. Inset shows the amplified local region around the state "Ts". (B) Free energy landscape calculated using the radius of gyration of all heavy atoms (R_g) and the fraction of native contacts. Both landscapes of (A) and (B) are calculated based on an ensemble of conformations with C_α-RMSD $\leqslant$ 0.45 nm. (C) Snapshots of structure taken along the folding and mis-folding processes with key residues colored in yellow. Only heavy atoms are shown for clarity. (For interpretation of color mentioned in this figure the reader is referred to the web version of the article.)

that Trp6-χ_2 leaves the "constrained" region. From Fig. 3A, we empirically derive L_{key} = 0.65 nm, which serves as a natural criterion for the collapse of the backbone of key residues. The second is a time for the backbone of key residues to loosen, as defined by $\Delta t_k = t_k - t_h$, where t_k is the FPT that $L_{key} > 0.70$ nm. Note that L_{key} = 0.70 nm represents the critical distance beyond which the backbone of key residues loosens significantly. To reduce noise, the temporal trajectories of both Trp6-χ_2 and L_{key} are filtered by a sliding window (length = 100 ps), and the calculation of the times is performed for all 40 trajectories, as shown in Fig. 4.

It is clear that Δt_c and Δt_k are strongly correlated, as shown in Fig. 4, with a correlation coefficient to be 0.98. However, we can infer that as Δt_c increases to be very large, the linear relation between the two parameters will deviate: in the extreme situation, Trp6-χ_2 is absolutely locked in the "constrained" region and so the loosening of backbone is dominated by spontaneous thermal fluctuation, in which the Δt_k will reach an equilibrium value denoted by Δt_k^∞. In particular, the quantity $\tau_f = 1/\Delta t_k^\infty$ is named as "spontaneous" reconfiguration rate. On the other hand, the loosening of backbone due to the torsion of Trp6-χ_2 will be called "side-chain-induced" (abbreviated by "induced") reconfiguration, and its rate is determined by $\tau_i = 1/\Delta t_k$, where Δt_k is far smaller than Δt_k^∞ and is specifically associated with Δt_c. The distribution of Δt_c displays an exponential tail, as shown in the inset of Fig. 4, which

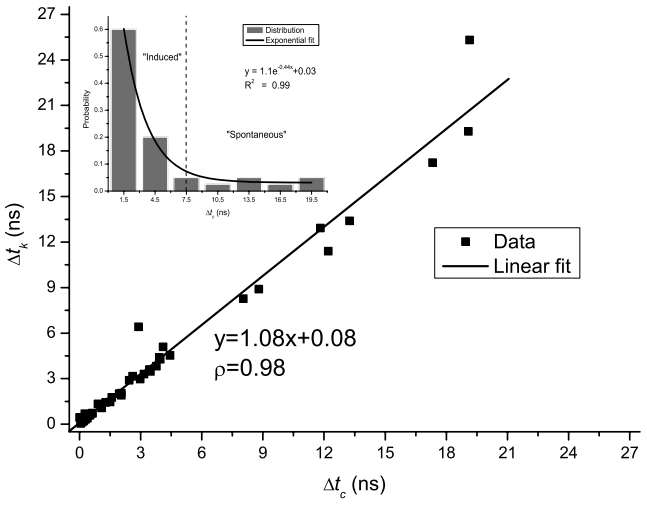

Fig. 4. Correlation between Δt_c and Δt_k. The definition of Δt_c and Δt_k can be found in the text. ρ is the correlation coefficient, and the data for all 40 trajectories are shown. Inset is the distribution of Δt_c (grey bar) with its first-order exponential fit (the black curve). Trajectories are classified into "induced" and "spontaneous" by Δt_c = 7.5 ns, which are used to study the rates of the corresponding two reconfiguration processes, respectively (see text).

indicates that in most cases the variation of Trp6-χ_2 and then the "induced" reconfiguration occur in very short time.

For a quantitative study of the rates of above two reconfiguration processes, we have classified trajectories into two categories as following: one is that $\Delta t_c < 7.5$ ns and represents an ensemble of the "induced" reconfiguration, while the other with $\Delta t_c \geqslant 7.5$ ns is used to approximate the "spontaneous" reconfiguration, and the criterion $\Delta t_c = 7.5$ ns is derived empirically from the inset of Fig. 4, where we find a clear separation of the pattern of distribution. In total, 8 "spontaneous" and 32 "induced" reconfiguration processes are discovered, and their reaction rates are $\tau_f^* = 1/\langle \Delta t_k \rangle_f = 1/14.59 \approx 0.07\ ns^{-1}$ and $\tau_i^* = 1/\langle \Delta t_k \rangle_i = 1/2.07 \approx 0.48\ ns^{-1}$, where $\langle\rangle_f$ and $\langle\rangle_i$ means taking ensemble average for "spontaneous" and "induced" processes, respectively. Note that τ_i^* is faster than τ_f^* by almost an order of magnitude.

The "spontaneous" reconfiguration is explicitly postulated in Kussell's "breathing" model [8]. In this model, the state of side-chain is constrained by the collapse of backbone and can relax only when backbone undergoes thermal fluctuation. Kussell's model predicts a very slow dynamics for the side-chain relaxation which resembles the processes that are observed in glass transition. However, their work based on a simplified model of side-chains, seems to lack sufficient atomic details to cover all forces properly, and therefore may miss important processes. Indeed, the present work shows that a possible faster reconfiguration may exist, in which the freely movable side-chains of specific key residues play a dominant role.

Conclusions

The concept of "key residues" was previously proposed to describe the formation of "key contacts" in specific transition states of folding as a critical step to trigger "downhill" backbone collapse [3,16]. Bendova-Biedermannoval et al. [17] have also proposed the term "key residue" in the study of stability of Trp-cage. The latter found that Trp6 plays a central role in stabilizing the native state of Trp-cage and so recognized it as the "key residue". In present work, we have proposed that there are a few key residues that are highly sensitive to the evolution of molecular configuration. Our "key-residue" hypothesis is more general; all-atom MD simulations of the folding of Trp-cage near the native state have identified four key residues (out of 20) whose parameters better describe the free energy landscape of the system. The new order parameters associated with key residues seem to be more efficient in differentiating among folding and mis-folding states, compared to the traditional parameters such as the radius of gyration and the fraction of native contacts. The significance of the new order parameters indicates the critical importance of the side-chain packing at sites of key residues.

Furthermore, the study of one of the key residues, Trp6, shows that there are two kinds of reconfiguration processes, side-chain-induced and spontaneous; the former has a faster reaction rate than the latter by almost an order of magnitude. The latter process was also observed in simplified models such as Kussell's model, for which the essential atomic details are missing. We find that these details are critical to recovering all proper forces and to represent a true folding process.

Compared to previous proposals, the key residues in our work are associated with the whole dynamical process of the protein and can offer a reduced key-residue-based phase plane analysis. This is of a general nature, since our work suggests a possible alternative way for coarse-grained description of the protein dynamics: only a few key residues are represented in sufficient atomic details while other residues are simplified. This method is now under development.

Acknowledgments

We acknowledge many useful discussions with M. Gao, H. Kang, H.Q. Zhu, Zhirong Liu, and the support by the National Natural Science Foundation of China (No. 10225210 and No. 30300071), and the National Basic Research Program of China (973 Program) under Grant No. 2003CB715905.

References

[1] E.I. Shakhnovich, Protein folding thermodynamics and dynamics: where physics, chemistry, and biology meet, Chem. Rev. 106 (2006) 1559–1588.
[2] K.A. Dill, Theory for the folding and stability of globular proteins, Biochemistry 24 (1985) 1501–1509.
[3] V.I. Abkevich, A.M. Gutin, E.I. Shakhnovich, Specific nucleus as the transition state for protein folding: evidence from the lattice model, Biochemistry 33 (1994) 10026–10036.
[4] F. Ding, N.V. Dokholyan, S.V. Buldyrev, H.E. Stanley, E.I. Shakhnovich, Direct molecular dynamics observation of protein folding transition state ensemble, Biophys. J. 83 (2002) 3525–3532.
[5] H. Li, R. Helling, C. Tang, N. Wingreen, Emergence of preferred structures in a simple model of protein folding, Science 273 (1996) 666–669.
[6] J. Liang, K.A. Dill, Are proteins well-packed?, Biophys J. 81 (2001) 751–766.
[7] E. Kussell, J. Shimada, E.I. Shakhnovich, Excluded volume in protein side-chain packing, J. Mol. Biol. 311 (2001) 183–193.
[8] E. Kussell, J. Shimada, E.I. Shakhnovich, Side-chain dynamics and protein folding, Proteins 52 (2003) 303–321.
[9] Y. Wei, W. Nadler, U.H. Hansmann, Backbone and side-chain ordering in a small protein, J. Chem. Phys. 128 (2008) 025105.
[10] K.A. Dill, Dominant forces in protein folding, Biochemistry 29 (1990) 7133–7155.
[11] J.W. Neidigh, R.M. Fesinmeyer, N.H. Andersen, Designing a 20-residue protein, Nat. Struct. Biol. 9 (2002) 425–430.
[12] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, The protein data bank, Nucleic Acids Res. 28 (2000) 235–242.
[13] W.L. Jorgensen, J. Tirado-Rives, The OPLS potential functions for proteins. Energy minimizations for crystals of cyclic peptides and crambin, J. Am. Chem. Soc. 110 (1988) 1657–1666.
[14] B. Hess, H. Bekker, H.J.C. Berendsen, J.G.E.M. Fraaije, LINCS: a linear constraint solver for molecular simulations, J. Comp. Chem. 18 (1997) 1463–1472.
[15] M. Ota, M. Ikeguchi, A. Kidera, Phylogeny of protein-folding trajectories reveals a unique pathway to native structure, Proc. Natl. Acad. Sci. USA 101 (2004) 17658–17663.
[16] M. Vendruscolo, E. Paci, C.M. Dobson, M. Karplus, Three key residues form a critical contact network in a protein folding transition state, Nature 409 (2001) 641–645.
[17] L. Bendova-Biedermannova, P. Hobza, J. Vondrasek, Identifying stabilizing key residues in proteins using interresidue interaction energy matrix, Proteins 72 (2008) 402–413.