Biochemical and Biophysical Research Communications 392 (2010) 95–99

Contents lists available at ScienceDirect

Biochemical and Biophysical Research Communications

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

Water dynamics clue to key residues in protein folding

Meng Gao [a], Huaiqiu Zhu [a,*], Xin-Qiu Yao [a,b], Zhen-Su She [a,*]

[a] State Key Laboratory for Turbulence and Complex Systems, and Department of Biomedical Engineering, and Center for Theoretical Biology, and Center for Protein Science, Peking University, Beijing 100871, China
[b] Department of Biophysics, Kyoto University, Sakyo Kyoto 606-8502, Japan

ARTICLE INFO

Article history:
Received 16 December 2009
Available online 7 January 2010

Keywords:
Key residues
Water dynamics
Protein folding
Trp-cage

ABSTRACT

A computational method independent of experimental protein structure information is proposed to recognize key residues in protein folding, from the study of hydration water dynamics. Based on all-atom molecular dynamics simulation, two key residues are recognized with distinct water dynamical behavior in a folding process of the Trp-cage protein. The identified key residues are shown to play an essential role in both 3D structure and hydrophobic-induced collapse. With observations on hydration water dynamics around key residues, a dynamical pathway of folding can be interpreted.

Introduction

With a series of experiments by residue mutation [1,2], it is now recognized that a few key residues in a protein sequence play key roles in the function, stability and folding of proteins. Meanwhile, many theoretical approaches have been proposed to investigate and even to predict key residues which are important in a structure of protein. Generally speaking, these theoretical studies can be summarized in two categories [3]. The first is based on sequential or structural alignments, under the argument that structurally or functionally important residues are highly conserved, which has been shown to depend on the known homolog information among proteins [3]. The second strategy aims at identifying key residues using contact or interaction energy evaluated by *ab initio* quantum chemical calculation [3], or studying on residues vibrational fluctuation pattern based on Gaussian network model or molecular dynamics (MD) simulation [3–5]. The latter mainly considers the key residues in folded protein structure stability analysis or in specific transition states of folding. However, there were few attempts to investigate key residues during a folding process, which are important to the understanding how proteins produce the correct 3D structure. A recent study by our group identified four key residues in the folding of the Trp-cage miniprotein (20 amino acid residues) using all-atom molecular dynamic simulation, which led to a proposal of key residue-dominated reconfiguration mechanism, in addition to spontaneous reconfiguration mechanism in protein folding [5]. However in that study, the experimental information, i.e. NMR structure, is needed as a reference to calculate a criteria function for key residues [5].

In most previous studies of key residues, a problem deserving more attention is the water–protein interaction. Until recently, the role of hydration water as an active component in the structure and the folding process of proteins has become accepted [6]. In fact, hydrophobic interactions have been shown to be the main driven force in protein folding [6]. More specifically, local hydration dynamics around key residues have cooperative effects with protein dynamics, and thus can provide information for protein structure and dynamics [7]. In a recent study by our group on wild type and mutant α-lytic protease differing by only one amino acid, we demonstrated that there are obvious distinctions in dynamic behaviors of hydration water [8]. Therefore, in such a hydrophobic-induced folding process, hydration dynamic behaviors is key to the understanding of water–protein interplay, and can be used to identify key residues.

In this paper, we present a computational method based on hydration water dynamics to recognize key residues in protein folding; the new method has the advantage of being independent of experimental protein structure information. Using data from all-atom MD simulations, two key residues are determined for the Trp-cage protein from their distinct water dynamical behaviors in the folding process. The key residues are shown to play an essential role in both structure and folding. With observations on hydration water dynamics around key residues, a dynamical pathway of folding can be interpreted.

Materials and methods

All-atom simulations. All analyses here are based on data of 4-μs folding trajectories from our previous work [5] for the Trp-cage system (PDB entry 1l2y), which was simulated using the GROMACS

* Corresponding authors. Fax: +86 10 6276 7261.
E-mail addresses: hqzhu@pku.edu.cn (H. Zhu), she@pku.edu.cn (Z.-S. She).

0006-291X/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.bbrc.2010.01.003

package [9] (Version 3.3) with the OPLS/AA force field and SPC water model. Starting from a partly unfolded configuration with randomly selected initial velocities, totally 40 simulations of the folding processes are performed in parallel at 282 K for 100 ns. The trajectories are saved every 10 ps, and then about 4,00,000 conformations are collected for analysis in this paper. There are 7 among 40 trajectories reaching the folded state.

Coarse-grained Gō-model simulations. To study the role of key residues in the folding kinetics of Trp-cage, coarse-grained simulations are performed with scaling of the interactions within a given residue pair or a contact group by a factor α. A similar strategy has been used in simulations of the coupled folding-binding process of intrinsically disordered proteins [10]. Herein, the Gō-model with coarse-grained C_α chain representation is used, and the potential model applied here is similar to that of the "without-solvation" model in [11]. Other parameters used here are set the same as in [11]. Langevin dynamics is used in dynamic simulations.

Results and discussion

Determining key residues without NMR information

To identify key residue, parameters such as RMSD and number of native contact have been used in previous studies. However, due to the need of a reference structure, the methods deploying those parameters are seriously dependent on the experimental information (e.g., NMR structure data). In the present study, we use the parameter *Rg*, radius of gyration (for all heavy atoms), to describe protein folding state, which can provide the fundamental information of protein collapse process without need for pre-known protein 3D structure. First of all, we classify all configurations collected in protein folding trajectory into 20 bins according to its *Rg* values. Then, taking *Rg* as a reaction coordinate, free energy may be calculated by

$$\Delta G_i = -k_B T \ln(Z_i) \quad (1)$$

where i is the index of 20 bins, Z_i is the probability of the system staying at the bin i, calculated by the percent of configurations in which *Rg* stays, and k_B is Boltzmann constant, T is Kelvin temperature, here taken to be 282 K.

Among the 20 bins characterized by *Rg*, the one where protein stayed with highest probability and hence with lowest free energy is defined as the reference set C^*. Herein the state C^* means a stable state and protein in the state C^* has collapsed structures. Then, the configurations with greater *Rg* are recognized as less collapsed, and are classified into four subsets by their *Rg* according to their free energy level. The subset classification schema is illustrated in Fig. 1A, and named from the most extended state to the most collapsed one with C_1, C_2, C_3, C_4, to C^*.

It has been argued that the hydration state of a protein shows the hydrophobic-driven protein collapse process [7]. Herein we define a parameter N_{hw}, the number of water molecules around a given residue side chain within a hydration shell (see below), to describe the hydration state of protein in folding process. By counting water molecules within the hydration shell around a given side chain along folding trajectory, then changes of N_{hw} can be measured for all residues in all trajectories. For Trp-cage in the current study, the hydration shell is defined to be 0.55-nm-thick with an irregular shape which covers all heavy atoms (non-hydrogen atoms) surface side chain, so that there will be no more than two layer of water molecules in the hydration shell. If oxygen atom of a water molecule is within 0.55 nm from any heavy atom of a given side chain, it will be regarded as one water molecule around this residue's side chain. During a protein collapse, some other residues take the places of water molecules which initially surround the center residues, and N_{hw} around the center residue then decreases. So, N_{hw} should be a sensitive measure to the residue which is located at the core center during a collapse process. In other words, the large fluctuation of N_{hw} will be an indicator of key residues in protein folding process, especially in hydrophobic-driven protein.

To measure the fluctuation of N_{hw} in a protein collapse process, we define the sensitivity, $Sn_i(k)$, of given residue k by

$$Sn_i(k) = \frac{\langle N_{hw}(k)\rangle_{C_i}}{\langle N_{hw}(k)\rangle_{C^*}} \quad (2)$$

here C_i represents the set C_1, C_2, C_3 or C_4. Apparently, $Sn_i(k)$ shows the relative value of N_{hw} around residue k when protein stays at a non-native state C_i relative to the near-native state C^*. Thus, when protein collapse from C_1 to C_4, some residues are packed more closely, and their *Sn* will decrease from a greater va-

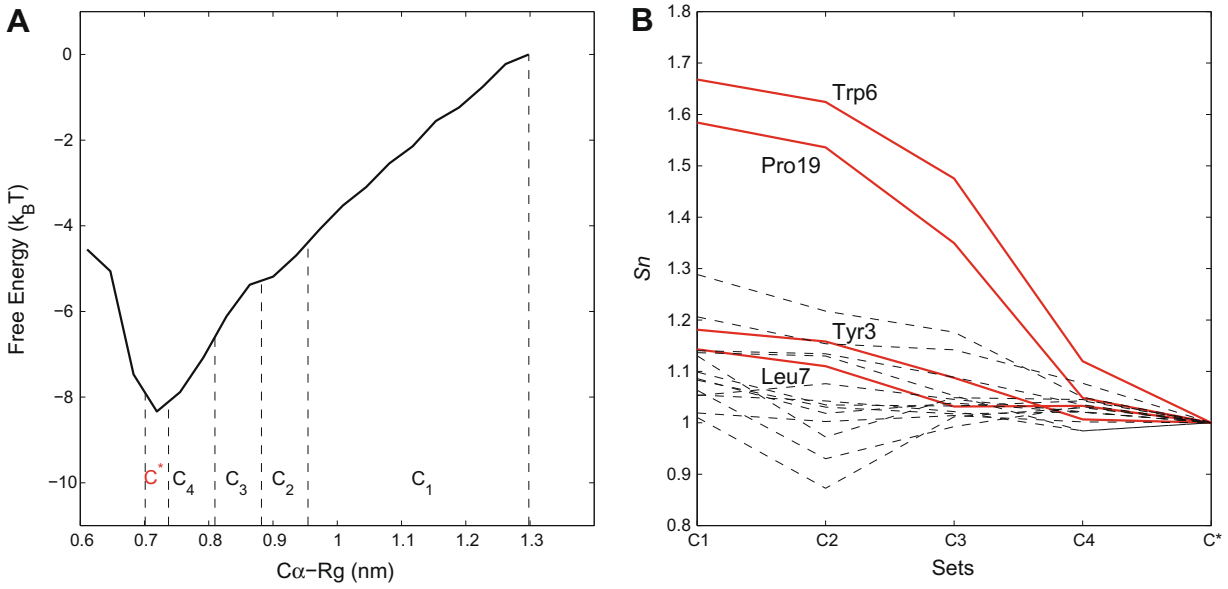

Fig. 1. (A) Free energy with Cα-*Rg* as a reaction coordinate. Here the maximum free energy is set to be 0 k_BT. C_1, C_2, C_3, C_4 is four defined subsets and C^* is the reference state, which is the bin with minimum free energy (see text). (B) Sensitivity scores (*Sn*) of 17 residues (Gly 10, 11, 15 are not included for being lack of side chains) in Trp-cage. The four key residues (Trp6, Pro19, Tyr3 and Leu7) identified in Ref. [5] are labeled and plotted as solid lines; *Sn* for other residues are plotted as dash-dot lines.