Computational and Theoretical Chemistry 978 (2011) 152–159

Contents lists available at SciVerse ScienceDirect

Computational and Theoretical Chemistry

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

Determination of thermal intermediate state ensemble of box 5 with restrained molecular dynamics simulations

Liling Zhao [a], Zhijun Liu [c], Zanxia Cao [a], Haiyan Liu [b], Jihua Wang [a,*]

[a] Shandong Provincial Key Laboratory of Functional Macromolecular Biophysics, Dezhou University, Dezhou, Shandong 253023, China
[b] School of Life Sciences and Hefei National Laboratory for Physical Sciences at the Microscale, University of Science and Technology of China (USTC), Hefei, Anhui 230027, China
[c] High Magnetic Field Laboratory, Hefei Institutes of Physical Science, Chinese Academy of Sciences, Hefei, Anhui 230031, China

ARTICLE INFO

Article history:
Received 27 July 2011
Received in revised form 6 October 2011
Accepted 6 October 2011
Available online 17 October 2011

Keywords:
Molecular dynamics simulation
Intermediate state prediction
High mobility group box domain
NMR experiment data
Free energy contour map
Cluster analysis

ABSTRACT

The thermal intermediate state of the high mobility group box 5 domain in human upstream binding factor was detected at 55 °C by nuclear magnetic resonance (NMR) experiments. For insufficient data, however, the tertiary structure of the intermediate state cannot be resolved as native state with the experimental techniques. To characterize the intermediate state ensemble, here we performed ensemble-averaged molecular dynamics simulations on the box 5 protein with 421 distance restraints derived from Nuclear Overhauser Enhancement and paramagnetic relaxation enhancement, as well as 122 dihedral angle restraints obtained from the program TALOS based on atom chemical shifts. The number of replicas was 48. The 60 ns simulation was completed for each replica. The total simulation time was up to 2.88 μs. The results indicated the intermediate state ensemble of box 5 was high heterogeneity and most secondary structures were formed. The N-terminal coil and helix 1 moved toward the C-terminal region; helix 3 was more stable and native-like than the other two helices; the hydrophobic core was not formed completely in the intermediate ensemble; and the L-shaped topology of the native conformation disappeared. In addition, some experimental inconsistencies were found, which could not be resolved in one conformation. In this study, the structural characteristics of box 5 thermal intermediate state ensemble were determined, which cannot be directly achieved through wet experiments. The findings of the current work are useful for the understanding of the protein folding mechanism. In our knowledge, this is the first report on the structural and thermal characters of the intermediate state ensemble of box 5.

1. Introduction

Protein intermediate state (IS) appears to play a crucial role in the folding and unfolding pathways of most globular proteins. Determining the conformational properties of IS is very important in understanding the mechanism of protein folding, aggregation, and misfolding. Nuclear magnetic resonance (NMR) spectroscopy is a powerful tool for studying protein characteristics in solution at atomic resolution, which can provide lots of information on protein structure, interactions, structural transformations, and dynamics et al. NMR can characterize protein denatured state more effectively than the other experimental techniques [1].

Because of the considerable increments in computer power, the molecular dynamics (MD) simulation method is an effective approach to investigating the characteristics of biomolecular systems [2–5]. Moreover, it is often used in the interpretation of NMR experiment data and refined protein structures [6,7]. Limited sampling in the higher dimensional conformational space is one of the most pressing problems in MD simulations. Getting the protein across the high energy barrier using traditional simulation techniques is difficult, preventing the spreading of the sample over all the relevant conformational spaces. With some restraints derived from NMR experiments, the sampling efficiency of MD simulations can be considerably improved. The restrained MD simulations are currently employed in studies on the unfolded state [8,9], intermediate state of protein [10] and some intrinsically disordered proteins [11,12].

NMR experimental values reflect average characteristics over a certain time frame (NMR time) and space (different conformations). Under native conditions, protein is in a relative equilibrium state

Abbreviations: HUBF, Human upstream binding factor; HMG, High mobility group; IS, Intermediate state; MD, Molecular dynamics; NMR, Nuclear magnetic resonance; NOE, Nuclear Overhauser Enhancement; PRE, Paramagnetic relaxation enhancement; $RMSD_{C\alpha}$, Root mean square deviations of Cα atom; RCs, Reaction coordinates; PDB, Protein Data Bank.

* Corresponding author. Tel.: +86 534 8985933; fax: +86 534 8985884.
E-mail addresses: zhaoll@sina.com (L. Zhao), zjliu@hmfl.ac.cn (Z. Liu), qiayilai@mail.ustc.edu.cn (Z. Cao), hyliu@ustc.edu.cn (H. Liu), jhwyh@yahoo.com.cn (J. Wang).

2210-271X/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.comptc.2011.10.004

and undergoes structural fluctuation around a well-defined conformation, which is important in performing its biological functions. Because of high stability, native structures can be achieved by MD simulations with all experiment values simultaneously satisfied [13,14]. However, under non-native conditions, protein rapidly transforms conformers at the NMR scale. Imposing all experimental values as restraints on one conformation is an unreasonable approach. Although conformations that satisfy all the restraints can be obtained, these conformations are excessively compact to reflect actual phenomena [15]. Structure ensembles are useful in characterizing proteins at non-native states [16]; the structure ensemble, and not only the single structure, should satisfy the restraints derived from NMR experiments.

Currently, three main approaches are adopted in calculating the structure ensemble, in which molecular dynamic simulation with experimental restraints is used. The first is using the time average over the simulation period by introducing a decay time factor [17–19]. The second approach pertains to assigning a weight factor to a pregenerated structure pool and obtaining a structure ensemble that agrees with the restraints, by which the unfolded state ensemble of drkN SH3 domain is characterized [20,21]. The third technique is called ensemble-averaged simulations. Restrained MD or Monte Carlo simulations are performed on multiple replicas in parallel. An ensemble-averaged simulation was employed to study the unfolded state of the Δ131Δ fragment of staphylococcal nuclease [22], the denatured state ensemble of ACBP [8,23] and the disordered state ensemble of α-Synuclein protein [11,12].

Human upstream binding factor (HUBF) is a nuclear transcription factor involved in transcription by RNA polymerase I. It contains six tandem-arranged high mobility group (HMG) box domains [24]; the solution structure of HMG box 5 is determined by NMR spectroscopy [14]. Box 5 shows a twisted L-shaped fold with a topology similar to that of the letter L. It consists of three α-helices: helix 1 (residues 9–25), helix 2 (residues 30–42), and helix 3 (residues 48–70). In another NMR experiment, the dynamic thermal IS ensemble of HMG box 5 was detected at 55 °C and confirmed by differential scanning calorimetry, circular dichroism spectroscopy, and fluorescence spectroscopy; however, the dynamic thermal IS ensemble could not be defined under insufficient data [25].

In this study, we performed ensemble-averaged MD simulations to study the IS ensemble of HMG box 5 at 328 K with 48 replicas. Simulations were performed up to 60 ns for each replica, and the total simulation time was up to 2.88 μs. The inter-atomic distances, which were derived from Nuclear Overhauser Enhancement (NOE) and paramagnetic relaxation enhancement (PRE), were used as restraints. In addition, 122 dihedral angle restraints were added to avoid artificial heterogeneity caused by distance nonlinear average r^{-6} in simulations [26,27]. The results indicate that the IS ensemble exhibited the most secondary structure as the native conformation and helix 3 behaved more stably than did helices 1 and 2. The angle between helices 1 and 2 increased and the hydrophobic core was not formed completely. The L-shaped topology of native conformation disappeared. These results are consistent with the experimental findings from circular dichroism spectroscopy and fluorescence spectroscopy [25]. This study is the first to depict the detailed structural characteristics of the IS ensemble of box 5 protein.

2. Methods

2.1. NMR experimental data

The production of HMG box 5 and the performance generated from the NMR experiments were described in detailed in our previous work [25] and the experiment data were collected at 55 °C. Backbone dihedral angles (φ and ψ) were obtained from the program TALOS [28] with $^{13}C\alpha$, $^{13}C\beta$, ^{13}C, ^{15}N, and $^{1}H\alpha$ chemical shifts as input. The dihedral angles with better prediction results (denoted as "good") were used as restraints during the simulations, and the total number of dihedral angle restraints was 122.

The $^{1}H-^{15}N$ NOE values were calculated as the peak intensity ratio of with-proton presaturation (I_{NOE}) to without-proton presaturation (I_{st}) [25]. The upper bound of the atom distance used as restraint was determined using this ratio. If the NOE value was larger than 100 or smaller than 50, the upper bound of the distance chosen was 3.5 or 5.5 Å, respectively, and the upper bound was set as 4.5 Å for the other NOE values. The lower distance bound was not assigned but the existence of repulsive force between atoms necessitated an atom distance larger than 1.8 Å. The number of distance restraints based on NOE was 234, including 60 intra-residue, 110 sequential, 41 medium-range, and 23 long-range constraints.

Because the wild-type box 5 is an intrinsically cysteine-free protein, three residues (N28, Y63, and E69) were selected for mutation to cysteine residue, and spin labeled for PRE experiments. Each mutant did not change the structural shape of the protein. The peak intensity ratio between the intensity of PRE (with spin label) and the reference (with inactive spin label) spectra, I_{PRE}/I_0, reflects the strength of interaction between the paramagnetic spin label and adjacent backbone protons (HN) [8]. This interaction is sensitive up to distances in the range 12–20 Å. The distance between two corresponding atoms was obtained based on the peak intensity ratio and the number was 187.

There were 122 dihedral angle restraints based on the chemical shift. At the same time, there were 421 distance restraints derived from NOE and PRE including 60 intra-residue, 110 sequential, 41 medium-range, and 210 long-range restraints.

2.2. Restraint MD simulations

To generate the initial conformations for ensemble-averaged simulations, one MD simulation was performed in vacuum at 600 K using the GROMOS96 software package [29,30] and the GROMOS 43B1 force field [29]. The simulation was initiated from the experimental NMR structure of box 5 (PDB entry 1L8Y, first model), and the velocity was generated by random seed. The dihedral angles that were denoted as "good" by TALOS in the helical structure were used as restraints. After 3 ns-long simulations, 48 conformations were obtained in succession at a time interval of 10 ps. These conformations were chosen as initial structures of the subsequent ensemble-averaged simulations.

In MD simulations with ensemble-averaged restraints, a number of non-interacting copies (or replicas) are simulated in parallel. Penalty functions based on restraints were added to the physical force field to enforce the agreement with the restraints, as follows:

$$V(t) = V_{phy}(t) + V_{dr}(t) + V_{\theta r}$$

where $V_{dr}(t)$ is the penalty function for distance restraints and is represented by

$$V_{dr}(t) = \sum_{m=1}^{N_{dr}} V_{dr}^{m}(t)$$

where the sum is over the N_{dr} distance restraints and $V_{dr}^{m}(t)$ is the function for the mth distance restraint; it is defined as [29]

$$V_{dr}^{m}(t) = \begin{cases} 0 & 0 \leqslant \langle r_m^{cal} \rangle \leqslant r_m^{\exp} \\ \frac{1}{2} N_{rep} K_d \left(\langle r_m^{cal} \rangle - r_m^{\exp} \right)^2 \cdot & r_m^{\exp} < \langle r_m^{cal} \rangle \leqslant r_m^{\exp} + \Delta r \\ N_{rep} K_d \left(\langle r_m^{cal} \rangle - r_m^{\exp} - \frac{\Delta r}{2} \right) \cdot \Delta r \cdot & r_m^{\exp} + \Delta r < \langle r_m^{cal} \rangle \end{cases} \quad (1)$$