Computational Biology and Chemistry 65 (2016) 148–153

Contents lists available at ScienceDirect

Computational Biology and Chemistry

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

Research Article

Side-chain dynamics analysis of KE07 series

Xin Geng[a], Jiaogen Zhou[b], Jihong Guan[a,*]

[a] Department of Computer Science and Technology, Tongji University, Shanghai 201804, China
[b] The Institute of Subtropical Agriculture, China Academy of Sciences, Changsha 410125, China

ARTICLE INFO

Article history:
Received 31 August 2016
Accepted 7 September 2016
Available online 15 October 2016

Keywords:
Molecular dynamics
Conformational entropy
Generalized order parameter
Protein design

ABSTRACT

The significant improvement of KE07 series in catalytic activities shows the great success of computational design approaches combined with directed evolution in protein design. Understanding the protein dynamics in the evolutionary optimization process of computationally designed enzyme will provide profound implication to study enzyme function and guide protein design. Here, side chain squared generalized order parameters and entropy of each protein are calculated using 50 ns molecular dynamics simulation data in both apo and bound states. Our results show a correlation between the increase of side chain motion amplitude and catalytic efficiency. By analyzing the relationship between these two values, we find side chain squared generalized order parameter is linearly related to side chain entropy, which indicates the computationally designed KE07 series have similar dynamics property with natural enzymes.

1. Introduction

Kemp eliminase KE07 is created by Röthlisberger et al. (2008) using de novo computational design methods (Baker, 2014) to catalyze the Kemp elimination reaction (Blomberg et al., 2013) which does not have a naturally occurring enzyme. To improve its catalytic efficiency (the K_{cat}/K_m value is about $12\,s^{-1}\,M^{-1}$), seven rounds of directed evolution, which generated a KE07 series with 7 variants, were carried out and resulted in a more than 200-fold increase of KE07's activity (the K_{cat}/K_m value is about $2600\,s^{-1}\,M^{-1}$) (Khersonsky et al., 2010). Inspired by natural selection in the natural world and artificial selection in human society, directed evolution in the laboratory is a highly efficient strategy to improve initial designs (Packer and Liu, 2015). The purpose of directed evolution in the laboratory is to mimic the process of biological evolution. Directed evolution does not require prior knowledge of structure–function relationship and it consists of iterative rounds of random mutagenesis and artificial selection aiming at achieving desired structures (Romero and Arnold, 2009). It starts with a diverse library of genes which later on is translated into a corresponding library of proteins. And then functional variants are screened or selected and replicated. They will be the starting points for the next round. The key active site residues of computational designed KE07 are Glu101 as the catalytic base, Trp50 interacting with the benzene ring and Lys222 as the hydrogen bond donor. While through 7 rounds of directed evolution, several mutations that are far away from these active site residues spatially and sequentially, e.g. Ile7 and Lys146, were made to increase the catalytic efficiency. Furthermore, the design process of KE07 did not analyze the molecular stability and dynamics. Here we mainly analyze the side chain dynamics of the KE07 series to discover the information of the structure–function relationship encoded in the directed evolution through the enzyme design process.

Dynamics plays a crucial role in enzyme catalysis and it provides critical information for the calculation of enzyme function and may guide protein design (Kohen, 2015). Molecular recognition of proteins with high-affinity interactions is the fundamental process underling enzymatic catalysis. Using NMR relaxation methods to study the side chain entropy of calmodulin, Frederick et al. (2007) showed that there is a linear relationship between the change in the overall binding energy and the change in the corresponding conformational entropy. While changes in conformational entropy cannot be reliably calculated from molecular structures, they are often represented by changes in conformational dynamics (Karplus et al., 1987). Molecular dynamics, which is a computer simulation method of protein motions by using Newton's equations of motions, can provide detailed information on the conformational dynamics. A popular way of capturing the essential character of the motion is to calculate the Lipari–Szabo squared generalized order parameter (Lipari and Szabo, 1982).

* Corresponding author.
E-mail addresses: gengxin_1986@163.com (X. Geng), zhoujg@isa.ac.cn (J. Zhou), jhguan@tongji.edu.cn (J. Guan).

http://dx.doi.org/10.1016/j.compbiolchem.2016.09.007
1476-9271/

Entropy is a measure of the freedom of a system to explore its available configurational space (Brady and Sharp, 1997). Side chain makes a main contribution to the reduction in conformational entropy that is synonymous with the binding of a protein to another protein or ligand. Side-chain dynamics also represent a major component of protein conformational entropy (Trbovic et al., 2009). Lipari and Szabo developed a "model-free" approach to capture the dynamical information (Igumenova et al., 2006). Assuming that the slower overall motion is independent of faster internal motions, the term "generalized order parameter" is generated from the mathematical approximation of the autocorrelation function for internal motion. The generalized order parameter, the limiting value of the autocorrelation function as well, represents the magnitude of the decay of the autocorrelation function due to internal motion.

The correlation of entropy (S) that is desired and Lipari–Szabo squared generalized order parameter (O^2) that can be measured has increasingly attracted wide attention and different models to illustrate the parametric relationship were proposed (Akke et al., 1993; Li et al., 1996; Yang and Kay, 1996; Li and Brüschweiler, 2009). Akke et al. (1993) showed the contribution to the free energy of binding can be obtained from generalized order parameters. Li et al. (1996) found that order parameters correspond to significant local entropies. Frederick et al. (2007) used generalized order parameters as a proxy for entropy calculation and found a surprising linear correlation with the change in total system entropy for binding. Through protein molecular dynamics simulations up to 600 ns, Li and Brüschweiler (2009) proposed the amino acid-specific relationships which constitute a dictionary for protein side-chain entropies from NMR order parameters. These relationships take a simple form, where S is calculated as a linear function of $1 - O^2$ or $log(1 - O^2)$ with only two parameters needed to fit. In their work, they listed different combination of N—H, C—H, and C—CH_3 bonds of side chains for different amino acids to make the most accurate estimation of side chain entropy by arithmetic average of O^2 values.

Even though the absolute entropy values estimated from O^2 may change based on the definition of models, the difference in the entropy values, a key factor to modulate the free energy of binding, are fairly insensitive to the model used. Therefore, based on previous works, for different amino acid type, we used the simplest linear function $y = ax + b$ to explore the relationship between side chain entropy S and O^2 as follows:

$$\frac{1}{M} S/k_B = a \cdot O^2 + b \tag{1}$$

where k_B is the Boltzmann constant, M denotes the number of side-chain dihedral angles and a and b are fit parameters. In our work, O^2 and S are calculated based on the data from molecular dynamics simulation of KE07 series under ligand-free ("apo") and ligand-bound ("bound") states. Then we try to investigate the changes of these two values and establish the linear model between them. Also we analyze the statistical changes along seven rounds of directed evolution.

2. Method

In order to get the parameters a and b of Eq. (1) by least squares fitting, firstly we need to calculate side chain O^2 and entropy. In this section, we'll describe the molecular dynamics simulations of KE07 series and then the calculation of side chain O^2 and entropy.

2.1. Molecular dynamics simulations

All the simulations ran on Teresa Head-Gordon's Lab servers at University of California, Berkeley.

We used the pmemd module in the AMBER package[1] for all of the molecular dynamics simulations. All the protein of KE07 series were modeled with the AMBER ff99SB protein force field (Hornak et al., 2006), and the TIP4P-Ew model (Horn et al., 2004) was used to describe the molecular solvent. The force field parameters were obtained from the generalized Amber force field protocol (Wang et al., 2004), with partial charges of each atom fitted to an HF/6-31G(d) electrostatic map using the RESP module in Amber. Each system was solvated in a rectangular water box with a buffer distance of 10Å between each wall and closest solute atom. After 200 ps of equilibration and sampling, the production phase of simulation then followed in the NPT ensemble(under constant pressure) for 50 ns.

We ran the simulation under NPT conditions using a weak barostat coupling at 1 bar and temperature of 300 K, using a 2 fs time step, with the long-range electrostatic interactions calculated using the Particle Mesh Ewald method and a cutoff of 12Å for real space electrostatics and LJ interactions. Trajectory snapshots were collected every 1 ps during the 50 ns simulation. The effects of overall motion during the simulation were removed by fitting each frame of the trajectory to the first frame. We used 5 ns for the time window to calculate the time autocorrelation function for each bond.

2.2. Side chain squared generalized order parameter O^2

The Lipari–Szabo model (Lipari and Szabo, 1982)interprets nuclear magnetic resonance relaxation experiments on macromolecules in solution. Assuming that the slower overall motion is isotropic and independent of faster internal motions, the time autocorrelation functions for internal motions of the bond unit vector orientations is defined as

$$C_l(t) = \langle P_2[\mu(0) \cdot \mu(t)] \rangle \tag{2}$$

where $P_2[x] = (3x^2 - 1)/2$, $\mu(t)$ is the bond unit vector at time t and the angle brackets $\langle\,\rangle$ denote ensemble averaging over the time window. Lipari and Szabo (1982) gave the simplest approximation to $C_l(t)$ with the form

$$C_l(t) = O^2 + (1 - O^2)e^{-t/\tau_e} \tag{3}$$

where O^2 is the squared generalized order parameter and τ_e is the effective correlation time.

For each bond, the time autocorrelation function with Eq. (2) was calculated directly by the ptraj module of Amber and O^2 was obtained by the least squares fitting of Eq. (3). We calculated the O^2 values for the N—H, C—H, and C—CH_3 bonds of each residue's side chain listed in Table 1 of Li and Brüschweiler (2009) and then took the arithmetic average as the O^2 value for each residue.

2.3. Side chain entropy S

The entropy of a system is given by the Boltzmann expression with terms of the probability of the system being in a particular configuration with an energy function (Brady and Sharp, 1997). While with the hardness of sufficiently exploring the conformational space and realistically modeling the potential energy function, we use the following two equations on basis of rotamer counting

[1] D.A. Case, R.M. Betz, D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E. Duke, T.J. Giese, H. Gohlke, A.W. Goetz, N. Homeyer, S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T.S. Lee, S. LeGrand, P. Li, C. Lin, T. Luchko, R. Luo, B. Madej, D. Mermelstein, K.M. Merz, G. Monard, H. Nguyen, H.T. Nguyen, I. Omelyan, A. Onufriev, D.R. Roe, A. Roitberg, C. Sagui, C.L. Simmerling, W.M. Botello-Smith, J. Swails, R.C. Walker, J. Wang, R.M. Wolf, X. Wu, L. Xiao and P.A. Kollman (2016), AMBER 2016, University of California, San Francisco.