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.

methods to estimate the side chain entropy. The side chain entropy for each residue is defined as:

$$S = -k_B \sum_{j=1}^{M} p_j \ln(p_j) \quad (4)$$

where k_B is the Boltzmann constant, M is the total number of non-clashing rotamers for each residue, and p_j is the probability of rotameric state j and it is calculated as

$$p_j = \frac{C(j)}{\sum_{i=1}^{m} C(i)} \quad (5)$$

where m is the total number of accessible rotameric states, $C(i)$ is the frequency of state i, and $C(j)$ is the frequency when rotameric state j is taken.

A backbone-dependent rotamer library created by Dunbrack and Cohen (1997) using a Bayesian statistical analysis of the conformations of side chains in proteins from the Protein Data Bank is widely used in homology modeling, protein folding simulation and the refinement of protein structures. Table 1 of their work lists all rotamers of all residue types with corresponding side chain χ angle ranges. This work makes entropy under our definition easy to calculate since the states of the side chain can be enumerated discretely from our molecular dynamics simulation data. For each residue, We use histogram binning to calculate the probabilities of each state in Eq. (5) and then the entropy of Eq. (4) according to the definition of all rotameric states of Dunbrack and Cohen (1997).

3. Results and discussion

Our goal is to use generalized order parameter to analyze the side chain dynamics of KE07 series and furthermore to investigate its relationship with entropy. In this section, we will show and discuss our results in 3 parts: the statistical distribution of side chain squared generalized order parameters, the link between side chain dynamics and catalytic efficiency and the linear relationship between side chain squared generalized order parameters and side chain entropy.

3.1. Distribution of side chain squared generalized order parameters

The backbone of designed KE07(PDB code:2RKXa) has 253 amino acids. As amino acid type ALA does not have side chain dihedral angles and GLY has a hydrogen substituent as its side chain, we did not calculate their O^2 values of side chain (they do not have side chain entropy values either). Eliminating the bonds whose time autocorrelation function did not converge within 50 ns of molecular simulation, finally we got 1613 average O^2 values for apo states and 1652 for bound states. Fig. 1 shows the histogram of the O^2 distributions of apo and bound states of KE07 series.

First of all, we compare the distribution of two states. We can see the frequencies of O^2 values in bound state are higher than those in apo state with big differences of 14, 22 and 23 for the ranges [0.1,0.2], [0.3,0.4] and [0.7,0.8]. O^2 is the term to interpret macromolecular internal dynamics by the "model-free" approach on the scale from 1, which means perfectly rigid, to 0, which means dynamic. So generally, we can conclude, comparing two states, the proteins are more rigid in bound state, even though for other ranges, the frequencies of O^2 in apo states is higher than in bound state by less than 10.

According to previous work on the distribution of the O^2, the O^2 values are clustered into three distinctive groups based on the correlation with the degree of motion freedom (Best et al., 2004; Trbovic et al., 2009), which represent highly restricted motion with values greater than 0.7, intermediate amplitude motion with values between 0.3 and 0.7 and large amplitude motion with values less than 0.3. Furthermore, Frederick et al. (2007) fitted the distribution histogram by 3-gaussion distributions centered on values of 0.35,0.58 and 0.78. Fig. 1 shows our data centered on values of 0.8–0.9 for both states and 0.3–0.4 for bound states.

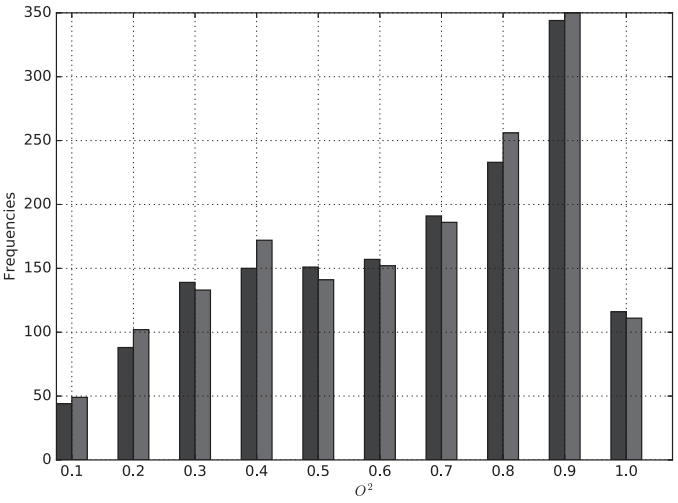

Fig. 1. Distributions of O^2 of KE07 series for apo (in blue) and bound (in green) states. The range [0,1] were divided into 10 intervals, e.g., "0.1" on horizontal axis represents range [0,0.1]. The value of vertical axis represents the counts of amino acids whose side chain O^2 values fell into each range. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

Three specific motional models were proposed in order to provide a detailed physical picture of the internal motions related with squared order parameters: diffusion-in-a-cone motional model, two-site jump motional model and combined diffusion-in-a-cone and two-site jump model (Jarymowycz and Stone, 2006). Our result that almost 1/4 of the values are between 0.8–0.9 showed the diffusion-in-a-cone motion in the ns-ps timescale of our molecular dynamics simulations. Previous NMR spin relaxation experiments showed the diffusion-in-a-cone, which means the bond vector is assumed to diffuse freely within but never move outside of a cone, was the most frequently employed method of calculating conformation entropy values from NMR-derived order parameters (Yang and Kay, 1996).

3.2. The link between side chain dynamics and catalytic efficiency

The significant increase of catalytic efficiency of KE07 variants generated by directed evolution shows the great success of evolutionary optimization of certain properties of natural proteins in the laboratory. Based on the listed kinetic parameters of KE07 variants in Röthlisberger et al. (2008), we plotted Fig. 2 to show the change of the catalytic efficiency with different directed evolution rounds. Meanwhile, Fig. 3 shows in the bound state, the change of average value of O^2 over all the residues and active site residues Trp50, Glu101, Lys222 and 13 residues mutated to create the designed Kemp eliminase active site (Khersonsky et al., 2010).

As mentioned before, the degree of motion freedom is defined into three kinds based on the O^2 values. In Fig. 3, two dashed lines with O^2 values of 0.3 and 0.7 are boundaries between different degree of motion freedom. The average O^2 value of all residues decreases as the catalytic efficiency increases with the directed evolution rounds. There is a big drop from 0.76 to 0.6 from the wild type to round 2, which means the overall motion changes form highly restricted to intermediate amplitude. And then average

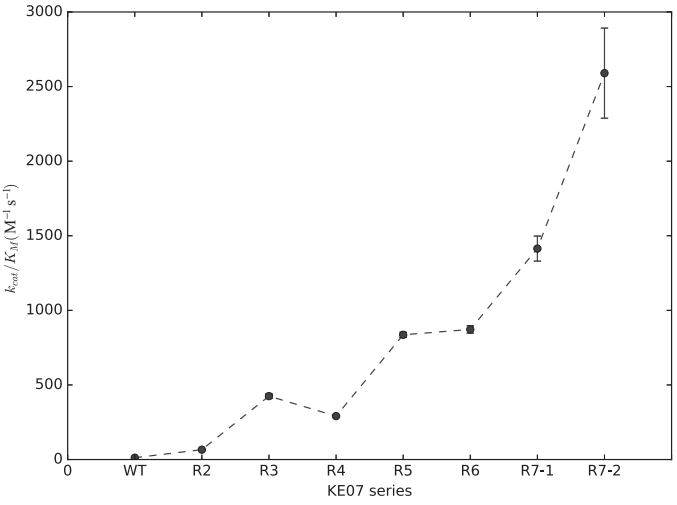

Fig. 2. The catalytic efficiency of KE07 series. Abbreviations used: WT represents wild type; R2, R3, R4, R5, R6 represent the variants from directed evolution round 2, 3, 4, 5, 6; R7-1 and R7-2 represent two conformers of round 7.

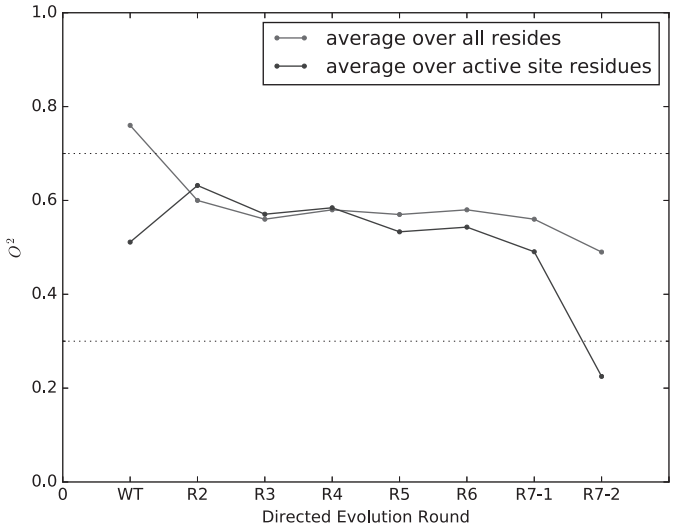

Fig. 3. Average value of O^2 with different directed evolution rounds in bound state.

O^2 gradually decreases to the minimum value of 0.49, which is still within the range of intermediate amplitude motion. The average O^2 value of active site residues goes up and down mildly while staying in the range of intermediate amplitude motion and then drops from 0.49 to the minimum value of 0.22 in the last two rounds, which means the motion changes into a large amplitude mode. The catalytic efficiency of round 4 is a local minimum value and this is consistent with our result that the average O^2 value of round 4 is a local maximum value.

We can conclude that, in seven rounds of directed evolution, the overall trend of average O^2 value is as opposite to that of catalytic efficiency. So our results, in terms of generalized order parameter, demonstrate the effects of the directed evolution on dynamics properties of the KE07 series, that gains of enzymatic functions

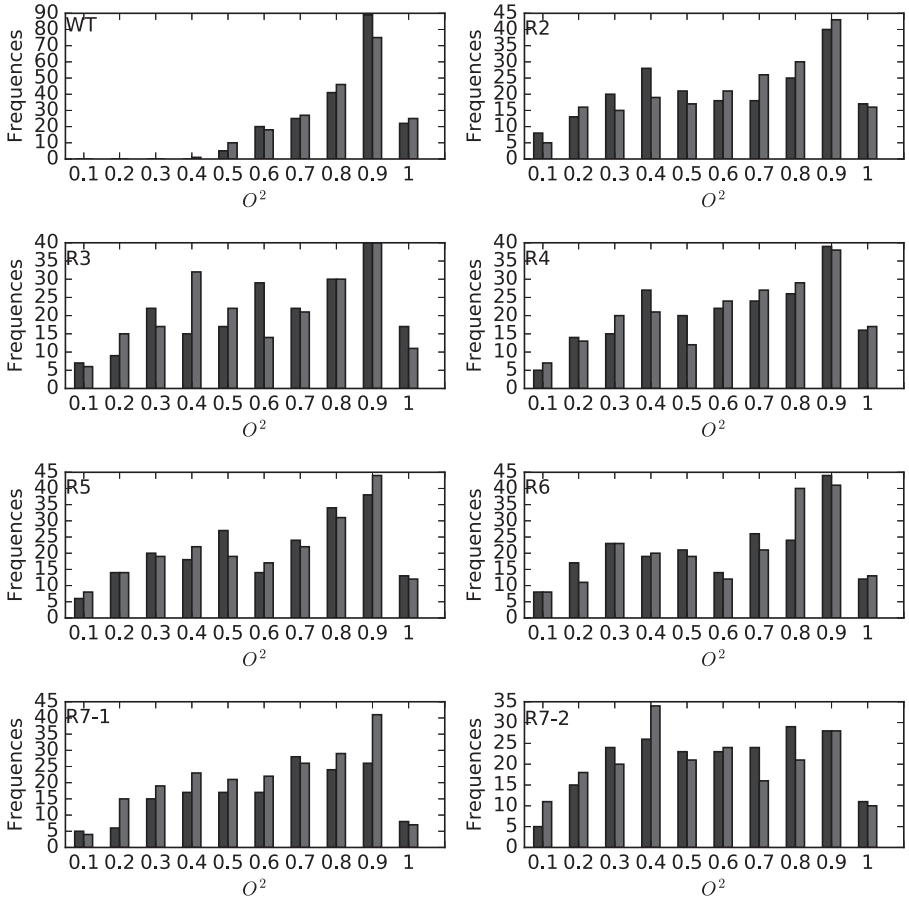

Fig. 4. Distribution of side chain O^2 values of KE07 series for apo (in blue) and bound (in green) states. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

Table 1
The fitting parameters of Eq. (1) for apo state.

AA	$-a/b$							
	KE07	R2	R3	R4	R5	R6	R7-1	R7-2
VAL	0.98/0.98	1.16/1.13	1.20/1.14	1.22/1.16	1.20/1.21	1.11/1.10	1.10/1.11	1.32/1.23
SER	1.20/1.09	1.17/1.10	1.21/1.12	1.00/1.00	0.98/1.01	1.05/1.06	1.17/1.08	1.10/1.09
THR	1.07/0.99	1.06/0.97	0.95/0.88	1.27/1.13	0.49/0.60	1.13/1.01	1.13/1.02	1.04/0.98
ILE	0.70/0.68	0.87/0.86	0.79/0.77	0.76/0.75	0.75/0.79	0.69/0.70	0.71/0.76	0.89/0.91
LEU	0.66/0.67	0.70/0.70	0.60/0.63	0.73/0.71	0.62/0.67	0.78/0.75	0.80/0.78	0.70/0.71
GLN	0.66/0.87	0.36/0.74	0.25/0.65	0.78/0.87	0.08/0.61	0.70/0.84	0.26/0.61	1.13/1.01
PHE	0.53/0.48	0.41/0.38	0.04/0.04	0.40/0.37	0.41/0.38	0.47/0.43	0.13/0.20	0.70/0.61
LYS	0.49/0.64	0.73/0.74	0.64/0.73	0.66/0.73	0.43/0.57	0.69/0.73	0.85/0.79	0.77/0.71
ARG	0.65/0.61	0.70/0.65	0.67/0.65	0.79/0.72	0.74/0.65	0.59/0.59	0.77/0.71	0.80/0.73
ASP	0.09/0.63	0.14/0.78	0.14/0.78	0.11/0.70	0.08/0.63	0.12/0.68	0.11/0.70	0.13/0.74
GLU	0.27/1.01	0.32/1.10	0.30/1.07	0.33/1.08	0.33/1.10	0.37/1.15	0.29/1.03	0.33/1.10

Table 2
The fitting parameters of Eq. (1) for bound state.

AA	$-a/b$							
	KE07	R2	R3	R4	R5	R6	R7-1	R7-2
VAL	1.12/1.04	0.32/0.48	1.15/1.11	1.19/1.15	0.92/0.90	1.23/1.14	1.31/1.21	0.14/0.34
SER	1.18/1.12	1.02/0.91	1.28/1.16	1.09/1.06	1.03/1.01	1.18/1.08	1.19/1.09	0.53/0.66
THR	0.91/0.89	0.46/0.66	0.90/0.89	1.33/1.23	0.15/0.32	1.25/1.11	1.11/1.02	0.43/0.47
ILE	0.84/0.80	0.37/0.50	0.32/0.53	0.46/0.56	1.03/0.96	0.59/0.67	0.74/0.75	0.45/0.48
LEU	0.61/0.64	0.36/0.51	0.69/0.68	0.85/0.83	0.47/0.54	0.64/0.67	0.59/0.64	0.32/0.40
GLN	0.55/0.72	0.70/0.84	0.31/0.82	0.46/0.78	0.48/1.28	0.01/0.51	0.73/0.92	0.56/0.40
PHE	0.15/0.17	0.06/0.08	0.16/0.16	0.48/0.43	0.18/0.22	0.17/0.19	0.46/0.44	0.39/0.38
LYS	0.72/0.75	0.79/0.77	0.70/0.73	0.64/0.70	0.47/0.57	0.68/0.71	0.80/0.77	0.49/0.54
ARG	0.70/0.67	0.32/0.44	0.53/0.59	0.69/0.67	0.65/0.61	0.65/0.62	0.60/0.59	0.23/0.33
ASP	0.12/0.73	0.07/0.54	0.12/0.71	0.13/0.70	0.16/0.80	0.10/0.65	0.10/0.64	0.11/0.68
GLU	0.27/1.04	0.25/0.94	0.31/1.10	0.30/1.05	0.32/1.10	0.43/1.23	0.40/1.19	0.23/0.89

cost the increased flexibility of the active-site region and reduced stability of the enzyme (Khersonsky et al., 2010).

We compared the changes in distributions of side chain O^2 value of KE07 series as shown in Fig. 4. As we can see, O^2 values of greater than 0.5 occur more often than those less than 0.5. All wild type O^2 values are greater than 0.3 and almost half fall into 0.8–0.9, while all the other variants of directed evolution have O^2 values spreading over all ranges. This may explain why the wild type has a lower catalytic efficiency compared with other variants. All the variants' distributions generally have peak values between 0.3–0.4 and 0.8–0.9, while the variant from round 3 (0.2–0.3, 0.5–0.6, 0.8–0.9 for apo state) and round 5 (0.2–0.3, 0.4–0.5, 0.8–0.9 for apo state) have 3 peak values. We also notice that the catalytic efficiency of variants from round 4 is lower than those form rounds 3 and 5, which is not consistent with the trend of increasing catalytic efficiency along the directed evolution.

As for the second variant from round 7, which has the best catalytic efficiency, the frequencies of side chain O^2 values from ranges 0.3–0.4 and 0.7–0.8 is almost same (difference of 2 for apo state). Especially for the bound state, range 0.3–0.4 has the highest frequency with 0.8–0.9 for all other variants. Furthermore, this variant's O^2 values have the most even distribution between 0.2 and 0.9.

3.3. The linear relationship between side chain O^2 and side chain entropy

For the apo and bound states of KE07 series, we fitted Eq. (1) for different residue types. In order to avoid the error caused by insufficient data, we excluded residue types appearing less than 10 times in the protein sequences, including Asn, Met, Tyr, Pro, His and Thr. We used least squares fitting in Matlab to get the parameters a and b of Eq. (1) and the results are listed in Tables 1 and 2 for apo and bound states respectively. All the linear models we got have a high coefficient of determination with $R^2 \geq 0.95$ and a low root mean squared error with $RMSE \sim 0.1$.

In the linear function expressed in Eq. (1), a is the slope parameter and b is the offset at the origin. As Tables 1 and 2 list, our fitted values of $-a$ and b are greater than 0, which is consistent with the values of fitted parameters of natural proteins in the work of Li and Brüschweiler (2009). However, In the work of Li and Brüschweiler (2009), $A+B$, the offset of the origin, takes values greater than 3 while all the b values of our work are less than 1.5, which shows the differences of the limits of entropy between different time scales(600 ns and 50 ns) of molecular dynamics simulations. As is known, O^2 represents the limitation of the bond vector autocorrelation function during global molecular tumbling, while $1-O^2$ characterizes the bond vector orientational disorder arising from molecular internal motion which happens more rapidly than the global tumbling. So from an experimental and theoretical point of view, Eq. (1) models the relationship between side chain O^2 and side chain entropy for KE07 series simply and accurately.

4. Conclusion

KE07 series are computationally designed enzymes for Kemp elimination reaction using directed evolution as the optimization method. In order to study the huge improvement in catalytic activity of KE07 series achieved by 7 round directed evolution, we use the generalized order parameter as a measure of side chain dynamics and investigate its relationship with side chain entropy. By analyzing the generalized order parameter values calculated on 50 ns molecular dynamics simulation data, we found changes in directed evolution showed that gains of enzymatic functions cost the increased flexibility of the active-site region and reduced stability of the enzyme. By using least squares fitting method, we found that side chain entropy is linearly correlated to the squared generalized order parameters for all the variants of KE07 series.

Acknowledgements

We thank Professor Teresa Head-Gordon from University of California, Berkeley for providing computing resources for molecular simulation, proposing using generalized order parameter and its relation with entropy to study dynamics of KE07 series and also thank her and Asmit Bhowmick for helpful discussions. This work was supported by National Natural Science Foundation of China (NSFC) under grant no. 61373036. Jihong Guan was also partially supported by the Program of Shanghai Subject Chief Scientist.

References

Akke, M., Brüschweiler, R., Palmer III, A.G., 1993. NMR order parameters and free energy: an analytical approach and its application to cooperative Ca^{2+} binding by calbindin d_{9k}. J. Am. Chem. Soc. 115 (21), 9832–9833.

Baker, D., 2014. Protein folding, structure prediction and design. Biochem. Soc. Trans. 42, 225–229.

Best, R.B., Clarke, J., Karplus, M., 2004. The origin of protein sidechain order parameter distributions. J. Am. Chem. Soc. 126 (25), 7734–7735.

Blomberg, R., Kries, H., Pinkas, D.M., Mittl, P.R.E., Grütter, M.G., Privett, H.K., Mayo, S.L., Hilvert, D., 2013. Precision is essential for efficient catalysis in an evolved Kemp eliminase. Nature 503 (7476), 418–421.

Brady, G.P., Sharp, K.A., 1997. Entropy in protein folding and in protein–protein interactions. Curr. Opin. Struct. Biol. 7 (2), 215–221.

Dunbrack, R.L., Cohen, F.E., 1997. Bayesian statistical analysis of protein side-chain rotamer preferences. Protein Sci. 6 (8), 1661–1681.

Frederick, K.K., Marlow, M.S., Valentine, K.G., Wand, A.J., 2007. Conformational entropy in molecular recognition by proteins. Nature 448 (7151), 325–329.

Horn, H.W., Swope, W.C., Pitera, J.W., Madura, J.D., Dick, T.J., Hura, G.L., Head-Gordon, T., 2004. Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. J. Chem. Phys. 120 (20), 9665–9678.

Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., Simmerling, C., 2006. Comparison of multiple Amber force fields and development of improved protein backbone parameters. Proteins 65, 712–725.

Igumenova, T.I., King Frederick, K., Wand, A.J., 2006. Characterization of the fast dynamics of protein amino acid side chains using NMR relaxation in solution. Chem. Rev. 106 (5), 1672–1699.

Jarymowycz, V., Stone, M., 2006. Fast time scale dynamics of protein backbones: NMR relaxation methods, applications, and functional consequences. Chem. Rev. 106, 1624–1671.

Karplus, M., Ichiye, T., Pettitt, B.M., 1987. Configurational entropy of native proteins. Biophys. J. 52 (6), 1083–1085.

Khersonsky, O., Röthlisberger, D., Dym, O., Albeck, S., Jackson, C.J., Baker, D., Tawfik, D.S., 2010. Evolutionary optimization of computationally designed enzymes: Kemp eliminases of the KE07 series. J. Mol. Biol. 396 (4), 1025–1042.

Kohen, A., 2015. Role of dynamics in enzyme catalysis: substantial versus semantic controversies. Acc. Chem. Res. 48 (2), 466–473.

Li, D.W., Brüschweiler, R., 2009. A dictionary for protein side-chain entropies from NMR order parameters. J. Am. Chem. Soc. 131 (21), 7226–7227.

Li, Z., Raychaudhuri, S., Wand, A.J., 1996. Insights into the local residual entropy of proteins provided by NMR relaxation. Protein Sci. 5 (12), 2647–2650.

Lipari, G., Szabo, A., 1982. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. J. Am. Chem. Soc. 104 (17), 4546–4559.

Packer, M.S., Liu, D.R., 2015. Methods for the directed evolution of proteins. Nat. Rev. Genet. 16 (7), 379–394.

Romero, P.A., Arnold, F.H., 2009. Exploring protein fitness landscapes by directed evolution. Nat. Rev. Mol. Cell Biol. 10 (12), 866–876.

Röthlisberger, D., Khersonsky, O., Wollacott, A.M., Jiang, L., DeChancie, J., Betker, J., Gallaher, J.L., Althoff, E.a., Zanghellini, A., Dym, O., Albeck, S., Houk, K.N., Tawfik, D.S., Baker, D., 2008. Kemp elimination catalysts by computational enzyme design. Nature 453 (7192), 190–195.

Trbovic, N., Cho, J.H., Abel, R., Friesner, R.A., Rance, M., Palmer, A.G., 2009. Protein side-chain dynamics and residual conformational entropy. J. Am. Chem. Soc. 131 (2), 615–622.

Wang, J.M., Wolf, R.M., Caldwell, J.W., Kollman, P.a., Case, D.a., 2004. Development and testing of a general Amber force field. J. Comput. Chem. 25 (9), 1157–1174.

Yang, D., Kay, L.E., 1996. Contributions to conformational entropy arising from bond vector fluctuations measured from nmr-derived order parameters: application to protein folding. J. Mol. Biol. 263 (2), 369–382.