Available online at www.sciencedirect.com
ScienceDirect
Progress in Natural Science 19 (2009) 137–144
Progress in Natural Science
www.elsevier.com/locate/pnsc

Review

Dynamics of DNA *in vitro* evolution

Xiaojing Yang [1], Xili Liu [1], Chunbo Lou, Qi Ouyang *

Center for Theoretical Biology, School of Physics, Peking University, Beijing 100871, China

Received 20 February 2008; received in revised form 5 May 2008; accepted 6 May 2008

Abstract

In vitro evolution has become a very important research area in recent years. From a practical point of view, it provides a powerful and reliable tool for engineering functional molecules (DNA, RNA or proteins) in reasonably short periods of time. From a theoretical point of view, since *in vitro* evolution is analogous to natural evolution in many respects, the study of the dynamic details of *in vitro* evolution may provide some instructive insights into the process of evolution. In this review, we summarize current theoretical and experimental studies, including several efforts made by our group, on the dynamics of DNA *in vitro* evolution.

Keywords: DNA-TF interaction; DNA *in vitro* evolution; Natural evolution; Crossover

1. Introduction

Next year will be the 200th anniversary of the birth of Charles Darwin and the 150th anniversary of the publication of his book "*on the Origin of Species by Means of Natural Selection or the Preservation of Favoured Races in the Struggle for Life*" [1]. The impact of Darwin's studies on the biological sciences can never be overstated. Charles Darwin provided us a general view to interpreting all biological forms of the world. In short, the Darwinian evolution process can be described as follows: there are inheritable variations of a population, and many more individuals are produced each generation than the capacity of the environment. Since individuals with adaptive characteristics are more likely to be selected to reproduce, over long periods of time, a population can become well adapted to a particular environment [1].

Forty years ago, Spiegelman and co-workers performed a study on the evolution of RNA molecules in a test tube that showed the possibility of carrying out the Darwinian evolution as a pure *in vitro* process [2]. Through iterative rounds of amplification, mutation, and selection, rare molecules with new functions were selected from an enormous collection of molecular possibilities. *In vitro* evolution provides a powerful and reliable tool for engineering functional molecules (DNA, RNA or proteins) in reasonably short periods of time, and it has become a highly fruitful area of investigation [3–8]. More importantly, since *in vitro* evolution is analogous to natural evolution in many respects, the study of the dynamic details of *in vitro* evolution can provide some instructive insights into the process of evolution.

The dynamic study of *in vitro* evolution, however, is presently still at an early stage. Although thousands of *in vitro* selection/evolution experiments have been performed to seek different types of targets, in most of them, only the terminal evolutionary pool was inspected for patterns. Some researchers have attempted to track the dynamics of fitness in the evolving population by measuring the catalytic activities of ribozymes and deoxyribozymes in the whole population during the *in vitro* process. However, the complexity of the correlation

* Corresponding author. Tel.: +86 10 62756943; fax: +86 10 62759041.
E-mail address: qi@pku.edu.cn (Q. Ouyang).
[1] These authors contributed equally to this work.

1002-0071/$ - see front matter © 2008 National Natural Science Foundation of China and Chinese Academy of Sciences. Published by Elsevier Limited and Science in China Press. All rights reserved.
doi:10.1016/j.pnsc.2008.05.029

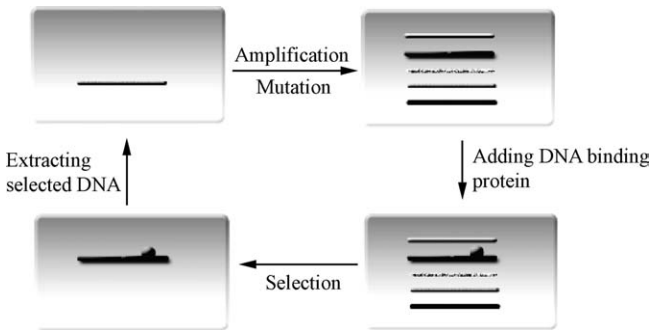

Fig. 1. A schematic view of the TF-DNA *in vitro* evolution system.

between the genotype (DNA/RNA sequence) and the phenotype (catalytic activity) prevented them from studying the dynamics of the evolution process in detail [9,10]. In comparison, the transcription factor (TF)-DNA system is a simpler system for directly determining the correlation between the genotype (DNA sequence) and phenotype (TF binding affinity). A general schematic view of TF-DNA *in vitro* evolution system is shown in Fig. 1.

The first experimental study of the dynamics of TF-DNA *in vitro* evolution was performed by Dubertret et al. in 2001 [11]. Some recent efforts have been made on the dynamics of the DNA *in vitro* evolution system, and several interesting dynamic features have been found. The present review summarizes both theoretical and experimental studies on the dynamics of DNA *in vitro* evolution so far, with a comparison of theoretical results and experimental results for each character of DNA *in vitro* evolution. In addition, future directions of *in vitro* evolution are also proposed.

2. Crossover phenomenon

2.1. Evidence of crossover phenomenon

In 2001, Dubertret et al. [11] started with a random pool of DNA sequences and monitored its ability to evolve into a subset of sequences with high affinity for the LacI. By monitoring the sequence distance to the native binding site, two separate groups of sequences were found to coexist during a given period of evolution: a group of sequences with good binding affinity and a group of random sequences. The authors concluded that instead of following a smooth process, *in vitro* evolution passed through a rapid transition point, which they called a crossover [11]. A crossover phenomenon was also observed in the stochastic simulation of the evolution of DNA in the DNA/Mnt-repressor system, and correlated with a sharp transition of the binding affinity of the entire population (Fig. 2b) [12].

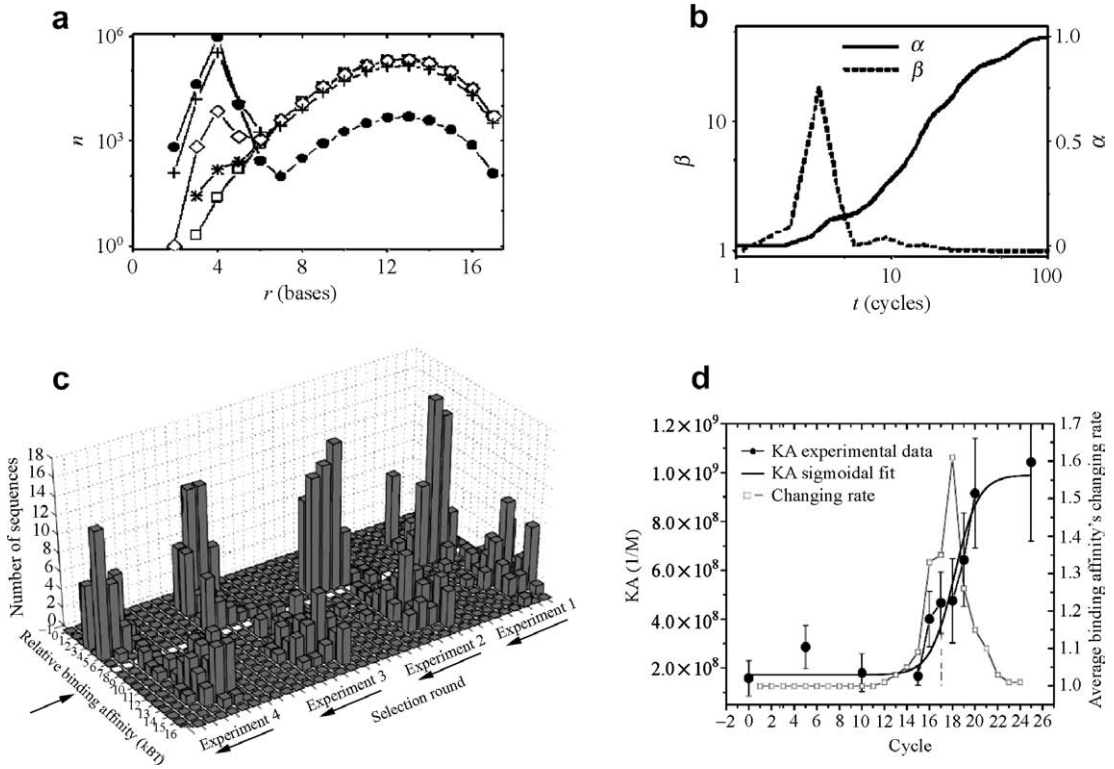

Fig. 2. Comparision of the crossover in simulations and experiments.[2] (a) Simulation: the time evolution of distance distribution $n(r)$ for typical simulation runs that exhibit the phenomenon of crossover (r represents the sequence difference to natural binding sequence for a given sequence); (b) simulation: the average binding affinity of the population (α) and the changing rate of the average binding ability (β) during the evolution process [12]; (c) experiment: distribution of the binding energy distance during the DNA *in vitro* experiment; (d) experiment: binding constant of the evolving population and the average binding affinity changing rate during the evolution process.

Dubertret's experiment was carried out under increasing selection pressure. Such increasing selection pressure would indicate that environmental conditions are always becoming more stringent and that the selection pressure is monotonically increasing during the evolution process. Since more often than not the environment in a natural system probably remains a near-constant during a period of time before changing to a new plateau, dynamic features of the DNA *in vitro* evolution process under a constant environment have been investigated recently.[2] Both the average binding affinity of the whole population and the distribution of the binding affinity of sequences in the evolving population were monitored in that experiment. A clear, ubiquitous and sharp crossover from a random population to an advantageous population was found, and a sharp transition of the binding affinity of the whole population was also observed experimentally. A comparison of simulation results and experimental results are shown in Fig. 2.

2.2. The mechanism of the crossover

DNA *in vitro* evolution processes show a crossover between two distinct sequence groups instead of a continuous improvement process. This suggests that the interaction between protein and DNA has two different states. Indeed, some studies on the structure of DNA-protein complexes have proven that the binding between DNA and protein has two different binding conformations: specific binding and non-specific binding [13–15]. Hence, if an energy difference exists between two distinct sequence groups during the evolution process, it is more likely that the energy difference between sequence-specific binding and sequence-non-specific binding exist.

In 2001, this hypothesis was first evaluated theoretically in the simulation of the Mnt-DNA *in vitro* evolution system [12]. The authors employed a real Mnt-repressor/DNA binding affinity matrix, and assumed that the binding energy of DNA to the Mnt-repressor consists of two parts, the sequence-specific energy and the sequence-non-specific energy. Two types of dynamics, crossover and non-crossover, were observed by adjusting the parameters of the energy difference between specific binding and non-specific binding (a_0). They concluded that, for the Mnt-DNA system, the crossover occurs only when $a_0 \leqslant 12k_BT$, while smooth evolution (non-crossover) occurs when $a_0 \geqslant 14k_BT$. If $a_0 \in [12k_BT, 14k_BT]$, the occurrence of the crossover strongly depends on the selection pressure.

This hypothesis was also evaluated experimentally.[2] First, it was proved that the interaction between the Mnt-repressor and DNA has two different energy states: sequence-specific binding and sequence-non-specific binding, and the binding energy difference between them was measured. Second, the energy difference between the two sequence groups during the evolution was monitored. As a result, the energy difference between the two sequence groups during the evolution in five independent experiments show a high agreement with the measurements of the energy difference between sequence-specific binding and non sequence-specific binding ($\sim 7k_BT$). Interestingly, the energy difference between the sequence-specific and sequence-nonsepecific binding observed in the experimental system ($\sim 7k_BT$) is exactly the energy difference required to observe a typical crossover in the simulation of the same system.

Combining the simulation results with the experimental results, it is very likely that the energy difference between two DNA/protein binding states, sequence-specific and non-specific, determines the crossover. Such an explanation of the crossover has been proposed (Fig. 3). If the contribution of each base pair in a DNA sequence to its protein binding affinity is independent and additive, the relative binding energy distribution for a random DNA pool should be a quasi-continuum, as shown by the gray lines in Fig. 3a. In this case, the selection process should also be continuous because the major selective force is the difference in the protein binding affinity. However, if there exists an energy boundary, beyond which the sequence variation does not influence the binding affinity, the relative binding energy distribution for a random DNA pool will be separated into two groups: a specific binding energy distribution and a non-specific binding distribution. The latter is like a delta function, as shown in Fig. 3b. During the evolution process, the peak of the first group moves towards the best binding sequence and its distribution becomes Gaussian-like, while the position of the second group remains almost constant, as shown in Fig. 3c.

If all the DNA sequences in the initial DNA pool lie in the specific binding regime, the crossover will not occur, as in the last case of Fig. 3c. The occurrence of the crossover requires a low energy boundary which can separate the population into two distinct groups in the binding affinity space, as in the first three cases in Fig. 3c. However, in practice, whether a crossover can be observed depends not only on the distribution of the initial DNA pool, but also on the applied selection pressure. If there exists a significant amount of specific binding sequences in the initial pool (such as the 3rd case of Fig. 3, maybe the 2nd case as well), and the selection pressure is high, most non-specific sequences might be removed at the first round of evolution, and a smooth evolution process will be observed instead of a crossover.

2.3. Timing of the crossover onset

Experimental results show that the crossover occurs after a long latent period during which there are no obvious changes in the population phenotype.[2] Different latent period times were found in independent evolution experiments under the same conditions. This has also been observed in the simulation of the *in vitro* evolution system

[2] Yang X J, Liu X L, Lou C B, et al. A case study on the dynamics of *in vitro* DNA evolution under constant selection pressure (J Mol. Evol. in press).

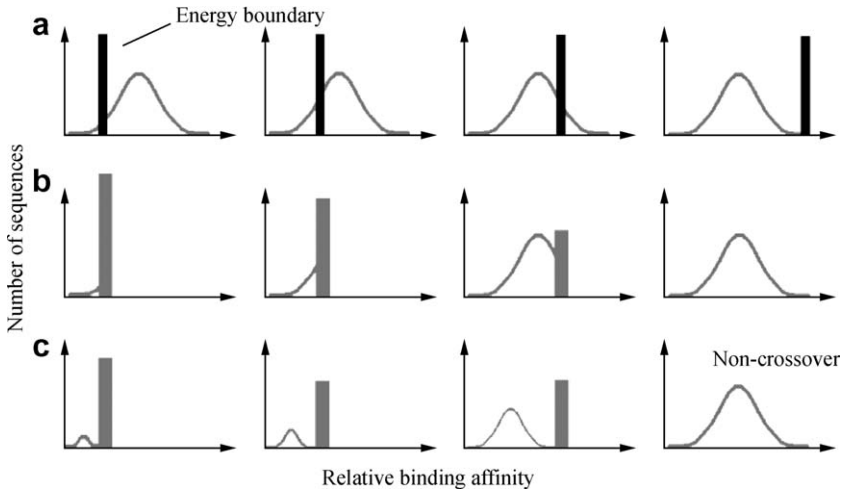

Fig. 3. The correlation between crossover and the energy boundary between specific and non-specific binding.[2] (a) The hypothetic relative binding affinity distribution for a random DNA pool (Grey curve); (b) the distribution of relative binding affinities for a random DNA pool with an energy boundary (black bar shown in (a); grey curves show the relative binding affinity distribution of specific sequences, while the grey bars show the relative binding affinity distribution of non-specific sequences); (c) a possible relative binding affinity distribution of the population at some later time points during the evolution process (each column shows different cases with different energy boundaries).

[12]. In general, the latent period time is required for specific binding sequences to emerge in the population by mutation. It will be affected by both the mutation rate, selection pressure, and the size of the population. In limited sampling experiments, it is also influenced by the number of samples.

3. Phase selection

Since a typical protein binding sequence only contains few base pairs and theoretically any continuous string of base pairs with required length can be the potential evolution targets in natural evolution, phase selection (which string will be selected) has also been inspected during the *in vitro* evolution process. In a recent Mnt-DNA *in vitro* evolution experiment, a 71-base long sequence, with 21 random base pairs in the middle and a 25-base long primer on both sides, was examined during evolution. Experimental results show that no evident phase could be identified at the beginning – the distribution of phases seemed random and most phases overlapped with primers (right or left). However, as the evolutionary cycles continued, phases converged into a few phases which have a little overlap with primers. The reason why most phases in which primer sequences were largely involved were lost in the evolution process might be that the random region can provide a larger potential space for further evolution (Fig. 4).

4. Quasi-equilibrium states

In 2003, based on a continuum mean-field model, Peng et al. studied analytically the dynamics of competitive DNA *in vitro* evolution. The authors concluded that interactions between mutations and the selection pressure can drive the system to an asymptotic equilibrium state where the population distribution centers on a sequence which can be far away from the best sequence that the protein binds (Fig. 5a) [16].

The quasi-equilibrium state was also found in the *in vitro* experiments under a modest constant selection pressure.[2] After the sequences with higher binding affinities became the dominant group in the population, the evolving pool was separated into two: one maintained in the original modest selection pressure environment, while the other was placed under a stronger selection pressure. The energy distribution of the population in the modest selection pressure remained constant in the following 10 rounds. In contrast, under a stronger selection pressure the second population distribution spread out and evolved towards the best sequence (Fig. 5b).

5. The correlation between the evolution speed and sequence diversity

In 1958, Fisher stated his "Fundamental Theorem of Natural Selection" in the form of: "The rate of increase in fitness of any organism at any time is equal to its genetic variance in fitness at that time" [17]. Fisher obtained his result on the basis of a continuous time model with logarithmic fitness. Assuming the fitness in Fisher's theorem is the binding affinity of DNA sequence, the Fisher theorem can be tested with the DNA *in vitro* evolution process.

Fisher's theorem was verified in DNA *in vitro* evolution process first by simulation in 2003, followed by further theoretical analysis under the assumption that the mutation rate is zero and the population size is infinity [12]. Then, Fisher's theorem was tested in a DNA *in vitro* evolution process experimentally.[2] A positive linear correlation between the changing rate of average binding affinity and the variance in binding affinity was also observed. Never-

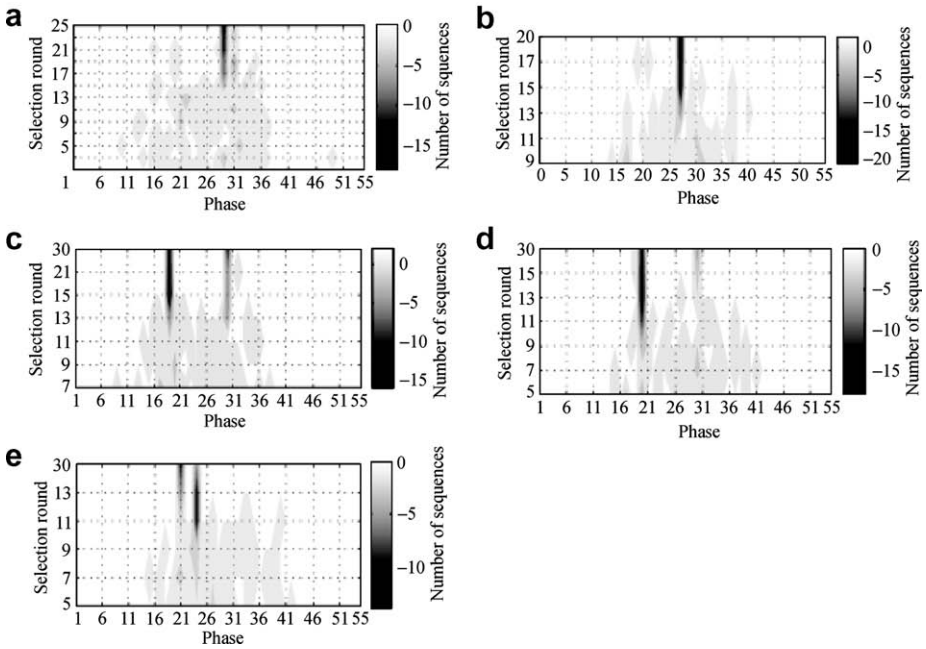

Fig. 4. Phase selection.[2] (a) Experiment 0; (b) experiment 1; (c) experiment 2; (d) experiment 3; (e) experiment 4.

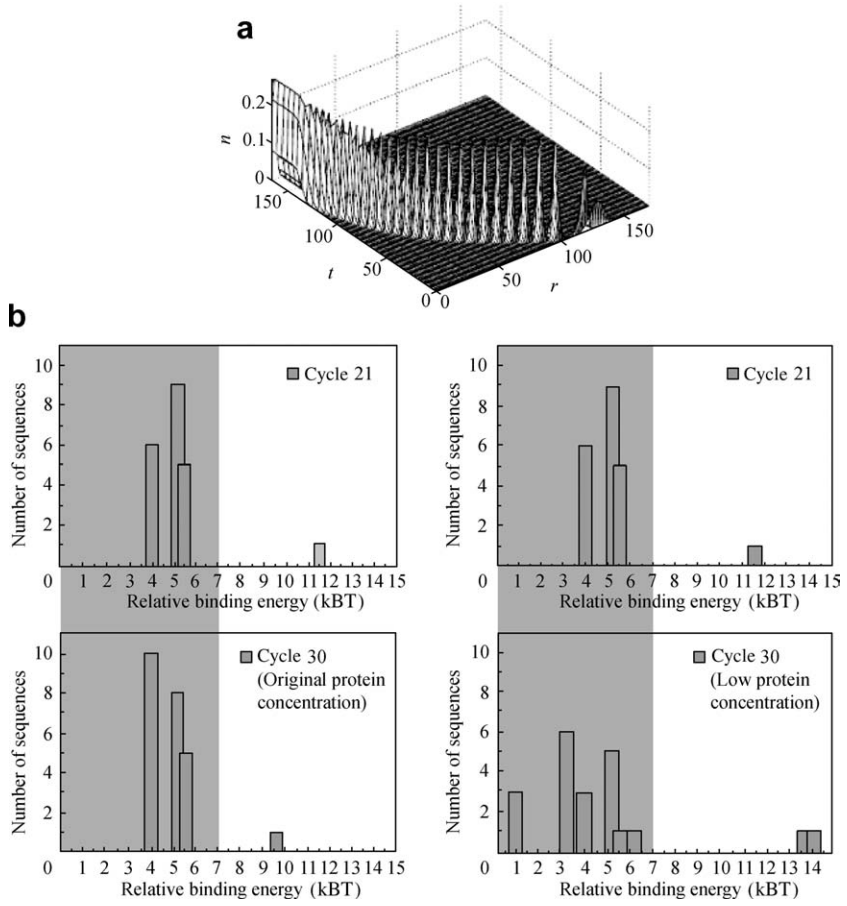

Fig. 5. Comparison of the quasi-equilibrium states in simulation and experiments.[2] (a) Simulation: the spatio-temporal trajectory of the mismatch distribution $n(r, t)$ according to the competitive evolution model (Ref. [17]); (b) experiment: evolution under different selection pressures. The charts on the left are the results of evolution with modest selection pressure (original protein concentration), while the charts on the right are with strong pressure (low protein concentration). Shaded boxes denote the binding affinity distribution of specific sequences. Twenty DNA fragments were chosen randomly at the given selection round for the binding energy distribution analysis. Note that the low protein concentration was applied at cycle 22 in this experiment.

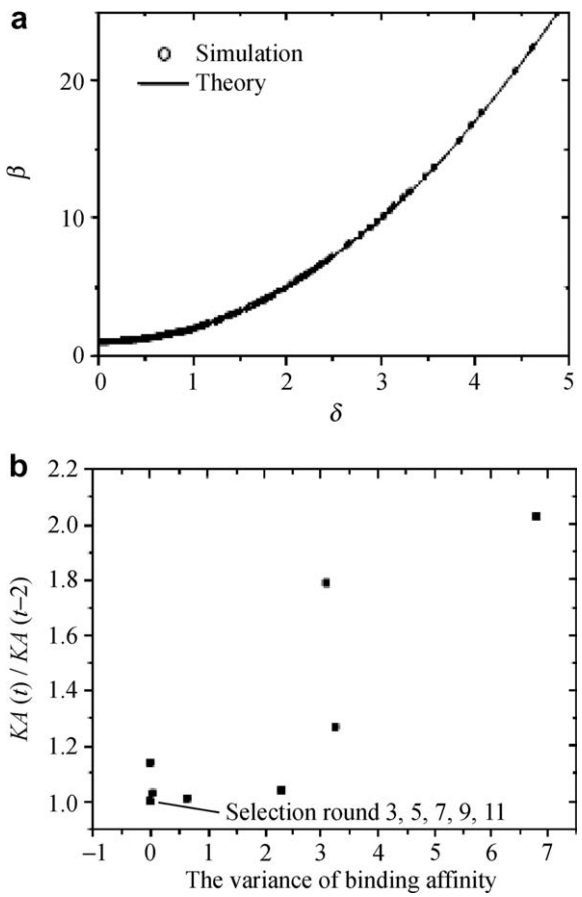

Fig. 6. The relationship between the changing rate of binding affinity and the variance of the binding affinity.[2] (a) Simulation: the dependence of β on the standard deviation δ of the selection probability that follows Fisher's theorem of natural selection (The simulation results (squares) agree very well with the function $\beta = 1 + \delta^2$ (solid line); parameters: $a_0 = 7k_BT$ [12]); (b) experiment: the relationship between the changing rate of binding affinity and the variance of the binding affinity for each odd cycle.

theless, due to the short-lasting population which had a large variance in fitness, there is insufficient data for a quantitative testing of the Fisher theorem experimentally. More independent evolution experiments under different selection pressures are required (Fig. 6).

6. Evolution trajectories

Evolution trajectories are always a charming issue of the natural evolution process. The evolution trajectories from different simulation runs were inspected in 2003 [12]. Results show that the evolution trajectories are drastically dispersed and no typical evolution passage exists during the evolution. However, the evolution trajectories of different simulation runs from identical initial random DNA populations extend less than that from different initial DNA pools, suggesting that the components of the initial population will partially affect the evolution trajectories.

In 2004, Kloster and Tang investigated the evolution trajectories from the given sequence (OT), which has six different base pairs from the natural binding site (WT). It was found that the WT sequence was produced through one of the minimum paths, if $N < \frac{1}{v_0}$, where N is the number of sequences in the initial pool, and v_0 denotes the mutation rate. The minimum path was defined as the path that only contains the six required mutations, and the number of minimum paths in this case is $6! = 720$. If N is pretty large, WT will be mostly produced through minimum paths, and if N is large enough, the contribution of each minimum path to the evolution is determinate [18].

7. A model for *in vitro* evolution

Based on the recent approaches in DNA *in vitro* evolution, experiments in particular, a simple model has been proposed,[2] and is shown in Fig. 7. The DNA *in vitro* evolution process is driven by two opposing forces: the selection force and the mutation force. While the mutation force depends only on the mutation rate, the selection force, which is based on the fitness of sequences, depends on both the diversity of the sequences and the binding affinities environment, which is the protein concentration in this case.

As the diversity of the population changes with time, the selective force changes as well. At the beginning, almost all sequences are non-specific, with about the same binding affinity to the protein and an even but narrow distribution of fitness. There is therefore no selection force but only mutation force acting on the system. This stage of the evolution is illustrated in Fig. 7a. In this period, the distribution of DNA sequences remains random and unchanged. At some points some specific binding sequences appear in the DNA pool by chance, so that the selection force steps into action. The selection force selects sequences based on the energy difference, resulting in a broad distribution of specific DNA sequences in the system, as shown in Fig. 7b. Under a given mutation rate, greater diversity of specific binding energy in the evolving population gives a larger selection force on the population, causing a larger changing rate of the average binding affinity (β). At the crossover point, both the selection force and specific binding diversity reach the maximum. The maximum changing rate of the average binding affinity is given as a peak in β (Fig. 7(c)). After a group of sequences with higher binding affinity becomes the dominant group, the diversity of sequence binding affinities decreases, and the evolving speed therefore slows down. The fitness of population reaches a plateau eventually, at which point the selection force and the mutation force are equal, as shown in Fig. 7d.

When the protein concentration decreases suddenly, the difference of fitness becomes more pronounced by the differences in binding affinities. The diversity of sequences' fitness increases, creating a larger selection force. This results in further evolution towards the best sequence, although the binding affinity of each sequence remains the same (Fig. 7e). As more and more sequences with lower binding affinities are removed from the population, the diversity of sequences' fitness decreases (Fig. 7f). The fitness of the pop-

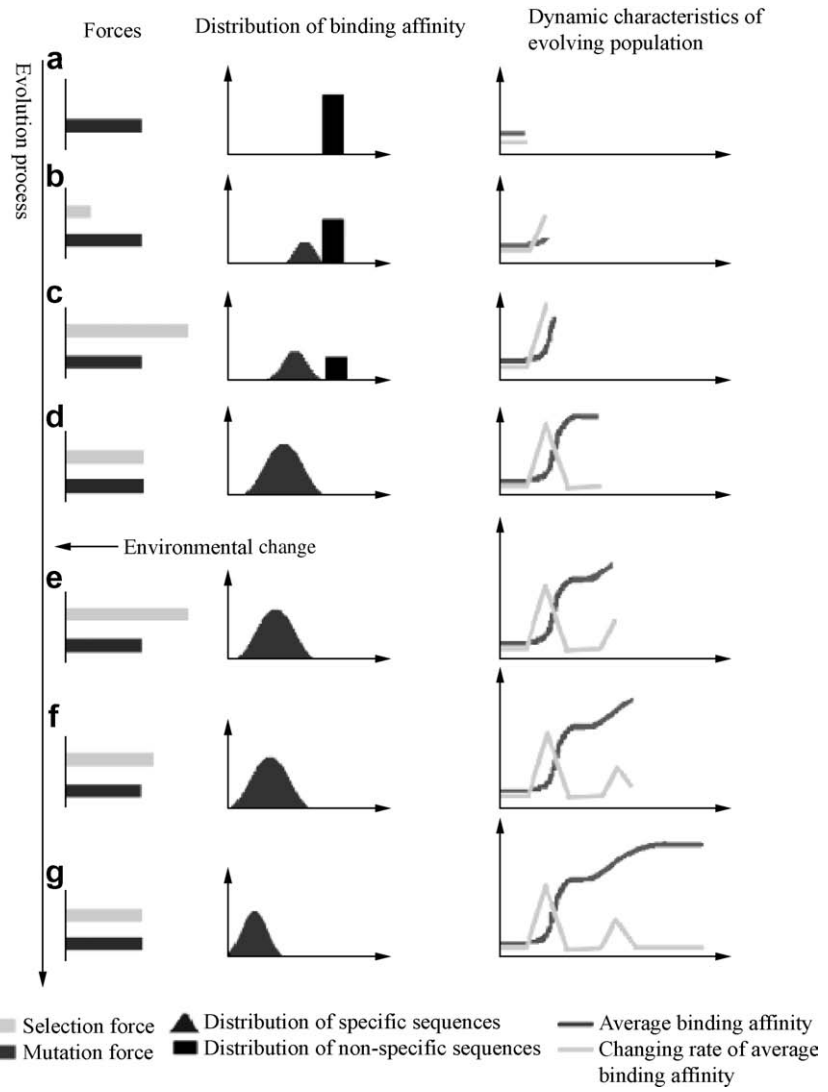

Fig. 7. A model for DNA *in vitro* evolution. (a)–(d), the dynamic response to the initial application of a constant selection pressure; (e) and (f), the dynamic response to the subsequent sudden increase in the selection pressure. The three different columns show the dynamic changes of forces, sequences' relative binding affinities distribution, and average binding affinity/changing rate of average binding affinity during the evolution process, respectively.

ulation eventually reaches a new plateau, where the selection force again equals the mutation force (Fig. 7g).

8. Perspectives

The interaction between protein and DNA is based on a DNA-protein complex energy landscape. Since the DNA *in vitro* evolution is based on a very simple interaction between DNA and protein, the DNA *in vitro* evolution landscape will be affected directly by the DNA-protein complex energy landscape. For example, two conformations of DNA-protein binding result in a crossover phenomenon during the evolution process[2] [12], while a smooth landscape leads to a smooth evolution process (non-crossover) [12]. Keeping this in mind, it would be very interesting to investigate the possibility of deducing some DNA-protein interaction information from the dynamics of the *in vitro* evolution process. Dubertret et al. identified some important base pairs for binding to lacI by the fixed order of each base pair during the evolution. Since the contribution of each base pair is not truly independent during the binding to a protein, it will be more reasonable to create some modules for DNA-protein binding by analyzing the correlation between different base pairs during the *in vitro* evolution process. Such work is currently being carried out in our group.

The notion of the dynamic study of *in vitro* evolution is still in its infancy, but this field is attracting more and more attention. We believe that as more dynamic features of *in vitro* evolution are discovered, both the evolution science and applied science fields will reap the benefits.

Acknowledgements

The authors thank Jennifer Hsiao for critical reading of the manuscript, and Jianguo Chen, Chao Tang, Terry Hwa, and Gary Stormo for their helpful suggestions.

References

[1] Darwin CR. On the origin of species by means of natural selection. London: John Murray; 1859.

[2] Bishop DH, Pace NR, Spiegelman S. The mechanism of replication: a novel polarity reversal in the in vitro synthesis of Q-beta-RNA and its complement. Proc Natl Acad Sci USA 1967;58(4):1790–7.

[3] Tuerk C, Gold L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. Science 1990;249:505–10.

[4] He Y, Stockley PG, Gold L. *In vitro* evolution of the DNA binding sites of Escherichia coli methionine repressor. MetJ J Mol Biol 1996;255(1):55–66.

[5] Hanes J, Pluckthun A. *In vitro* selection and evolution functional proteins by using ribosome display. Proc Natl Acad Sci USA 1997;94:4937–42.

[6] Keefe AD, Szostak JW. Functional proteins from a random-sequence library. Nature 2001;410:715–8.

[7] Matsuura T, Yomo T. *In vitro* evolution of proteins. J Biosci Bioeng 2006;101(6):449–56.

[8] Joyce GF. Directed evolution of nucleic acid enzymes. Annu Rev Biochem 2004;73:791–836.

[9] Bartel DP, Szostak JW. Isolation of new ribozymes from a large pool of random sequences. Science 1993;261:1411–8.

[10] Schlosser K, Li Y. Diverse evolutionary trajectories characterize a community of RNA-cleaving deoxyribozymes: a case study into the population dynamics of *in vitro* selection. J Mol Evol 2005;61:192–206.

[11] Dubertret B, Liu SM, Ouyang Q, et al. Dynamics of DNA-protein interaction deduced from *in vitro* DNA evolution. Phys Rev Lett 2001;86:6022–5.

[12] Yang YF, Wang HL, Ouyang Q. Dynamics of DNA *in vitro* evolution with Mnt-repressor: simulations and analysis. Phys Rev E 2003;68:0319031.

[13] von Hippel PH. Completing the view of transcriptional regulation. Science 2004;305:350–2.

[14] Spolar RS, Record MT. Coupling of local folding to site-specific binding of proteins to DNA. Science 1994;263:777–84.

[15] Lesser DR, Kurpiewski MR, Jen-Jacobson L. The energetic basis of specificity in the Eco RI endonuclease – DNA interaction. Science 1990;250:776–86.

[16] Peng W, Gerland U, Hwa T, et al. Dynamics of competitive evolution on a smooth landscape. Phys Rev Lett 2003;90:0881031.

[17] Fisher RA. The genetical theory of natural selection. Oxford: Oxford University Press; 1999.

[18] Kloster M, Tang C. Simulation and analysis of *in vitro* DNA evolution. Phys Rev Lett 2004;92(3):038101.