Chapter 42
The Mathematical Modeling and Computational Simulation for Error-Prone PCR

Lixin Luo
South China University of Technology, China

Fang Zhu
South China Universityof Technology, China

Si Deng
South China University of Technology, China

ABSTRACT

Many enzymes have been widely used in industrial production, for they have higher catalytic efficiency and catalytic specificity than the traditional catalysts. Therefore, the performance of enzymes has attracted wide attention. However, due to various factors, enzymes often cannot show their greatest catalytic efficiency and the strongest catalytic ability in industrial production. In order to improve the enzyme activity and specificity, people become increasingly interested in the transformation and modification of existing enzymes. For the structure modification of proteinase, this chapter introduces a computational method for modelling error-prone PCR. Error-prone PCR is a DNA replication process that intentionally introduces copying errors by imposing mutagenic reaction condition. We then conclude about the mathematical principle of error-prone PCR which may be applied to the quantitative analysis of directed evolution in future studies.

INTRODUCTION

The successful application of enzymes as industrial biocatalysts requires the availability of suitable enzymes with high activity, specificity and stability in process conditions. However, naturally procreant enzymes are often not optimized to meet these requirements(Zhao et al., 2002). Using the method of directed evolution, we can not only produce useful biocatalyst for the organic chemistry domain, but also improve the properties of the biocatalyst and even create biocatalysts that possess novel catalytic activities and properties (Raillard et al., 2001). Directed

DOI: 10.4018/978-1-4666-3604-0.ch042

evolution has been used increasingly in academic and industrial laboratories to modify and improve important biocatalysts (Arnold & Volkov, 1999). It is a fast way that simulates Darwinian evolution in the test tube to improve protein with a certain purpose (Zhao et al., 2002). Relative to the traditional "rational design", this method is called "irrational design". Error-prone PCR protocols, which use a low fidelity replication step to introduce random point mutations into DNA sequences at each round of amplification, were used in early directed evolution experiments (Moore & Arnold, 1996; Zaccolo & Gherardi, 1999).

Error-prone PCR is a DNA replication process that introduces copying errors in PCR process by changing reaction conditions. The key to the error-prone PCR is to control the DNA mutation frequency (Leung et al., 1989). If the DNA mutation frequency is too high, the vast majority of the procreant enzymes will lose their activity; if the mutation frequency is too low, the amount of wild-type enzymes will be too high and the diversity of the sample will be too low. The observed mutational frequency resulted in 0.25-20 mutations per 1000 base pairs in the error-prone PCR. The ideal rate of basic group displacement and the optimal error-prone PCR condition depend on the length of the target DNA fragment, the running time of PCR and the mutational frequency ω (Moore & Arnold, 1996).

MODELING FOR ERROR-PRONE PCR

The top priority of mutagenic PCR is to introduce various types of mutations in an unbiased form rather than to achieve a high overall level of amplification (Cadwell & Joyce, 1994). As in the regular PCR, the first step is the denaturing which the double-stranded DNA is separated into two single strands by heating; the second step is the annealing which is the primer binds to the complementary single-strand DNA; the third step is extension which the template sequence is extended by DNA polymerase. As non-complementary nucleotides can bind to the extended-chain, mutation occurs in the third step (Gregory & Costas, 2000). The error rate of *Taq* polymerase is the highest of the known thermostable DNA polymerases, in the range of 0.1×10^{-4} to 2×10^{-4} per nucleotide per pass of the polymerase, and depending on reaction conditions (Leung et al., 1989). It is important to control these highly variable copying errors for obtaining "useful" mutations and excluding "useless" mutations (Gregory & Costas, 2000). Simulation technique has become extremely important in almost every aspect of scientific and engineering endeavor(Neim, 1995). Simulation is experimentation with models(Korn & Wait, 1978).Therefore, we introduce computational method model into error-prone PCR and make a conclusion about mathematic law of error-prone PCR, then it can play a guiding role in the analysis.

In the proposed model, mutations will occur during the extending process and every mutation can be considered as a contrary event to the others. Let ω represent different mutation rates and a single mutation rate M_{ij} stands for the mutative probability from nucleotide i to nucleotide j.

$$\omega = \begin{bmatrix} M_{AA} & M_{AT} & M_{AC} & M_{AG} \\ M_{TA} & M_{TT} & M_{TC} & M_{TG} \\ M_{CA} & M_{CT} & M_{CC} & M_{CG} \\ M_{GA} & M_{GT} & M_{GC} & M_{GG} \end{bmatrix} \quad (1)$$

These values depend on the experimental conditions. ω can be used to describe mutation rate ω^n after n extension steps. When n=0, M_{ij}^n =0 (but if i=j, M_{ij}^n =1); when n=1, $M_{ij}^n = M_{ij}$; when $n \geq 1$, $M_{ij}^n = \sum_{k=A,T,C,G} M_{kj} M_{ik}^{n-1}$ (Gregory & Costas, 2000).

The Computational Method: