

EFFICIENCY OF POLYMERASE CHAIN REACTION PROCESSES: A STOCHASTIC MODEL

Arjang Hassibi

California Institute of Technology

Masoud Sharif

Boston University

ABSTRACT

A stochastic model for the efficiency of polymerase chain reaction (PCR) processes is presented. The model is based on the assumption that the i 'th nucleotide incorporation on the DNA template takes τ_i seconds, where τ_i has an exponential distribution with mean $1/\lambda_i$. In this paper, given λ_i that can be obtained from the primer-template sequence, temperature profile, enzyme rate, and other assay conditions [3], we calculate the efficiency of a multi-step PCR process using the distribution of the summation of non-identical exponential distributions.

1. INTRODUCTION

The typical goal of Polymerase Chain Reaction (PCR) processes is to replicate, and exponentially grow double-stranded DNA template molecules by using a set of primers (i.e., short single-stranded DNA fragments), which act as initiation sites for DNA polymerase enzyme. The efficiency of a PCR process is accordingly defined as the probability of the completion of the replication in each individual cycle of the process [1]. It has been shown that PCR efficiency can not only be used to evaluate the expected exponential growth of the template molecules, but also their probability distribution. The statistics of the PCR process, evaluated using the efficiency parameter, have also been further exploited to derive optimal estimators for PCR assays [2].

While the implementation of efficiency is a powerful approach to analyze and design PCR assays, currently there is no analytical model to find its value. All constructed models are based on assay-specific empirical data. To address such impediment, in this paper, we present a general analytical model to evaluate PCR efficiency.

2. DESCRIPTION OF THE MODEL

The efficiency of the PCR process is an intricate function of the assay characteristics, which include the enzyme type, template and primers sequence and concentration, temperature profile, and the reaction buffer ionic strength. Our method to analytically assess the efficiency is to first decouple and statistically model the underlying processes of PCR into hybridization incidents and enzyme-initiated polymerization.

In the PCR reaction mix, different DNA strands can interact and possibly hybridize to one another, forming double-stranded DNA molecules. The probability of finding a certain hybridized complex essentially depends on the concentration of the strands, base matching between them, ionic strength of the reaction buffer, and the temperature. For two strands s_i and s_j , the probability

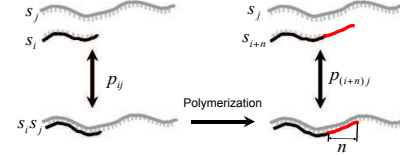


Fig. 1. Fundamental steps of the polymerization process in PCR where the primer-template probabilistic attachment may result in primer extension by the enzyme.

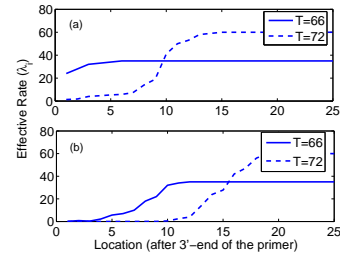


Fig. 2. Effective rate of polymerization for (a) primer-template perfect match, and (b) primer-template single mismatch at the 3'-end. Solid and dashed curves are for $T = 66^\circ\text{C}$ and $T = 72^\circ\text{C}$, respectively.

of finding an s_i template with an attached s_j molecule (denoted by p_{ij}), can be calculated using the Gibbs free energy change ΔG_{ij} , temperature T , and the concentration of s_j and s_i [3].

For the primer extension (polymerization), the enzyme first needs to attach to the $s_i \cdot s_j$ complex (s_i and s_j are the primer and template respectively as shown in Fig. 1). Next, s_i which acts as the primer gets extended with a polymerization rate of λ_i by the enzyme where λ_i is a function of temperature and enzyme type. It is imperative to realize that since $s_i \cdot s_j$ complex is an unstable system, the effective λ of the process decreases by any detachment incident. However, since we know that the probability of finding a primer on the template is p_{ij} , we can obtain the effective rate of incorporation with possible detachment at the i 'th incorporation λ_i , as λp_{ij} . In this model λ_i increases as a function of i , since longer DNA strands have less probability of detachment. In other words as i gets larger, p_{ij} tends to one and λ_i converges to λ . In Fig. 2 we have shown for two different sets of primer-template how λ 's behave. In both cases λ_i 's are converging to nominal λ as the primer extends and $s_i \cdot s_j$ becomes more stable.

In what follows, we obtain the distribution of the number of incorporations in a PCR cycle which would readily lead to the efficiency of the process.

3. EFFICIENCY FOR THE TWO-STEP PCR

In the two step PCR (which is the most basic amplification procedure used in PCR assays), we have one denaturing phase with high temperature and one polymerization phase with lower temperature. In our model and during the polymerization phase of duration t , we assume that the i 'th incorporation takes τ_i seconds where τ_i has exponential distribution with mean $1/\lambda_i$. From a biophysical point of view, this is a valid assumption since nucleotide incorporation in PCR process is very close to a memoryless process (which is based on the Poisson model for biochemical binding).

It is clear that m 'th incorporation would take $S_m = \sum_{i=1}^m \tau_i$ secs. Therefore the efficiency of PCR (i.e., complete synthesis of the template) is equal to $\Pr(S_n < t)$ where n is the size of the template strand. In other words, we may define the variable k_t which is the number of incorporations completed in t secs (i.e., $S_{k_t} \leq t$ and $S_{k_t+1} > t$). Thus the efficiency will be equal to $\Pr(k_t \geq n)$. In order to find the distribution of k_t , we write,

$$\Pr(k_t = m) = \int_0^t \int_{t-x}^{\infty} f_{S_m}(x) \lambda_{m+1} e^{-\lambda_{m+1}y} dy dx$$

where $f_{S_m}(x)$ is the probability distribution function of the random variable S_m .

As S_m is the summation of non-identical independent exponentially distributed random variables, the distribution can be readily obtained using the moment generating function of S_m as $E(e^{-sS_m}) = \prod_{i=1}^m E(e^{-s\tau_i}) = \prod_{i=1}^m \frac{\lambda_i}{\lambda_i + s}$. Since eventually λ_i is converging to λ as the extension length increases, we define m_{eq} 'th step to be the first step that its rate converges to λ , i.e., $\lambda_{m_{eq}} = \lambda$.

Case 1: $m \leq m_{eq}$: Using partial fraction expansion, we get

$$E(e^{-sS_m}) = \sum_{i=1}^m \frac{\alpha_{i,m}}{\lambda_i + s} \Rightarrow f_{S_m}(x) = \sum_{i=1}^m \alpha_{i,m} e^{-\lambda_i x}. \quad (1)$$

We can obtain the distribution of k_t when $1 \leq k_t < m_{eq}$ as,

$$\Pr(k_t = m) = e^{-\lambda_{m+1}t} \sum_{i=1}^m \frac{\alpha_{i,m}}{\lambda_{m+1} - \lambda_i} \left(e^{(\lambda_{m+1} - \lambda_i)t} - 1 \right)$$

It is clear that $\Pr(k_t = 0) = e^{-\lambda_1 t}$ and a similar expression can be found for $\Pr(k_t = m_{eq})$ that we omit for brevity.

Case 2: $m > m_{eq}$: As m increases, computing the partial fraction expansion in (1) becomes difficult and numerically unstable due to the fact that the fraction has a root of order $m - m_{eq}$. In order to avoid this problem, we can divide S_m into two parts as $S_m = \sum_{i=1}^m \tau_i = \sum_{i=1}^{m_{eq}} \tau_i + \sum_{i=m_{eq}+1}^m \tau_i = A_1 + A_2$. Here the distribution of A_1 is equal to $\sum_{i=1}^{m_{eq}} \alpha_{i,m_{eq}} e^{-\lambda_i x}$ and the distribution of A_2 is $\chi^2(2r)$ where $r = m - m_{eq} - 1$. Therefore,

$$\begin{aligned} \Pr(k_t = m) &= \Pr(A_1 + A_2 \leq t, \tau_{m+1} > t - A_1 - A_2) \\ &= \alpha_{m_{eq}, m_{eq}} \frac{\lambda^r}{(r+1)!} e^{-\lambda t} t^{r+1} \\ &+ \sum_{i=1}^{m_{eq}} \frac{\alpha_{i,m_{eq}} \lambda^r}{\lambda - \lambda_i} \left(e^{-\lambda_i t} \frac{\gamma((\lambda - \lambda_i)t, r)}{(\lambda - \lambda_i)^r} - e^{-\lambda t} \frac{t^r}{r!} \right) \end{aligned}$$

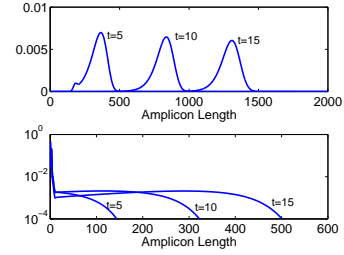


Fig. 3. The distribution of the polymerization length of the primer (amplicon) for (a) primer-template perfect match, and (b) primer-template single mismatch at the 3'-end.

where α 's are defined in (1) and $\gamma(x, m)$ is the incomplete gamma function defined as $\gamma(x, r) = \int_x^{\infty} \frac{t^{r-1}}{(r-1)!} e^{-t} dt$.

It is clear that computation of the distribution of k_t even for large m is quite easy and only requires the computation of one integral (i.e., incomplete gamma function).

4. MULTI-STEP PCR AND SIMULATIONS

In the previous subsection, we obtained the distribution of k_t in polymerization phase of a two-step PCR. If we have different temperature steps for polymerization (e.g., three step PCR where we have annealing and extension phase), the number of incorporation is just the sum of the number of incorporations in different steps, the distribution of k_t in the multi-step case is thus equal to the convolution of the distribution of k_t in different steps. Therefore, the result can be extended to PCR assays with multi-step temperature profile.

In our simulations, we considered a two-step PCR process, where the polymerization steps takes t seconds. In Fig. 3, we have plotted the distribution of k_t (number of incorporation in t) for different value of t and for the two cases of primer-template perfect match and primer-template single mismatch at the end of 3'-end.

It is interesting to observe that the probability distribution function of k_t for primer-template perfect match case has two local maximums (see Fig. 3), one close to zero and the other one is roughly about λt that corresponds to the likelihood that the incorporation passes the first few extremely slow steps and moves forward with the steady state λ .

5. REFERENCES

- [1] G. Stolovitzky and G. Cecchi, "Efficiency of DNA replication in the polymerase chain reaction," *Proceedings of National Academy of Science*, vol. 93, pp. 12947-12952, 1996.
- [2] N. Lalam, "Estimation in the setting of population-size-dependent branching processes: Application to quantitative polymerase chain reaction," *Ph.D. dissertation, University Paris*, 2005.
- [3] A. Hassibi and H. Kakavand and T. H. Lee, "A Stochastic model and simulation algorithm for polymerase chain reaction systems" *Proc. of Workshop on Genomics Signal Proc. and Stat.*, 2004.