Sequence-dependent biophysical modeling of DNA amplification

Karthikeyan Marimuthu[1], Chaoran Jing[2] and Raj Chakrabarti[1, 2*]

1. Department of Chemical Engineering and Center for Advanced Process Decision-making, Carnegie Mellon University, Pittsburgh, PA 15213
2. Division of Fundamental Research, PMC Advanced Technology, Mt. Laurel, NJ 08054

Abstract

A theoretical framework for prediction of the dynamic evolution of chemical species in DNA amplification reactions, for any specified sequence and operating conditions, is reported. Using the Polymerase Chain Reaction (PCR) as an example, we developed a sequence- and temperature-dependent kinetic model for DNA amplification using first principles biophysical modeling of DNA hybridization and polymerization. We compare this kinetic model with prior PCR models and discuss the features of our model that are essential for quantitative prediction of DNA amplification efficiency for arbitrary sequences and operating conditions. Using this model, the kinetics of PCR are analyzed. The ability of the model to distinguish between the dynamic evolution of distinct DNA sequences in DNA amplification reactions is demonstrated. The kinetic model is solved for a typical PCR temperature protocol to motivate the need for optimization of the dynamic operating conditions of DNA amplification reactions. It is shown that amplification efficiency is affected by dynamic processes that are not accurately represented in simplified models of DNA amplification that are the basis of conventional temperature cycling protocols. Based on this analysis, a modified temperature protocol that improves the PCR efficiency is suggested. Use of this sequence-dependent kinetic model in a control theoretic framework to determine the optimal dynamic operating conditions of DNA amplification reactions, for any specified amplification objective, is discussed.

Keywords: Polymerase Chain Reactions, Optimal Control, Kinetics, PCR efficiency

---

* To whom correspondence should be addressed – Email: raj@pmc-group.com, rajc@andrew.cmu.edu

1. **Introduction**

DNA amplification is the process of geometric growth of the number of double-stranded DNA (dsDNA) molecules in solution through repeated replication of single-stranded DNA (ssDNA) templates. Due to the universal need to amplify DNA for applications ranging from molecular cloning to DNA sequencing, such methods have arguably become the central technology of modern molecular biology. The polymerase chain reaction (PCR), the most common DNA amplification reaction, is a cyclic amplification process that can produce millions of copies of double-stranded DNA molecules starting from a single molecule. The traditional three-step PCR reaction cycle consists of (i) dsDNA denaturation, (ii) oligonucleotide primer annealing to the resulting ssDNAs, and (iii) polymerase-mediated extension steps to produce two dsDNA molecules. This cycle is repeated 20-30 times, resulting in geometric growth of the number of DNA molecules. The base of the exponent for geometric growth is termed the amplification efficiency of a cycle.

Despite the fact that the notion of thermal cycling is based on a dynamic picture of DNA amplification (1), there are currently no models of DNA amplification kinetics that are capable of predicting the evolution of reaction products for general sequences and operating conditions. Without such a model, the optimal temperature cycling protocol for the reaction – which is sequence specific - cannot be computed, and reductions in cycle efficiency (either through decreased reaction yield or specificity compared to the theoretical maximum values) can occur. Due to geometric growth, reductions in the cycle efficiency can result in dramatically diminished efficiency of the overall reaction, and substantial efforts have hence been dedicated to improving the efficiency of DNA amplification reactions (2,3).

In the absence of predictive models for DNA amplification, the operating conditions for PCR reactions are typically selected based on qualitative analysis of their kinetics and thermodynamics, given the desired amplification objective. Over the past two decades, many variants of DNA amplification have been invented based on the notions of DNA denaturation, annealing and polymerization, each tailored to a particular amplification objective. Each such reaction (which is typically assigned its own acronym) is based on a temperature cycling

protocol determined through analysis of reaction thermodynamics and a qualitative analysis of kinetics. A simple example of a temperature cycling protocol that involves modifications to the conventional prescription is the use of two-step PCR cycles (4), wherein annealing and extension occur simultaneously at a properly chosen temperature.

A general approach to kinetic modeling of DNA amplification has applications to the design of new types of amplification reactions, in addition to enhancement of existing reactions. In the language of systems engineering, the selection of the optimal trajectory of a manipulated input variable such as temperature is referred to as dynamic optimization or optimal control (5). This paper is concerned with the establishment of a foundation for the dynamic optimization of DNA amplification reactions, which can be used for the automated computation (rather than qualitative selection) of temperature cycling protocols. To date, quantitative sequence-dependent modeling of DNA amplification has been largely restricted to the thermodynamics of the reaction. Prior reports of kinetic models for PCR have proven inadequate for the purposes of dynamic PCR optimization. For example, Rychlik et al (6) developed an empirical equation to determine an optimal annealing temperature that maximizes the final DNA concentration. Using a probabilistic PCR kinetic model, Stolovitzky and Cecchi (7) developed a method to calculate the cycle efficiency for PCR quantification. Velikanov and Kapral (8) proposed a Markov process approach to optimize the extension step of PCR. Yang et al (9) discussed the effect of annealing temperature on the concentrations of different targets in a multiplex reaction and gave the temperature vs. concentration profile for all the targets. Though the above-developed approaches predict the PCR efficiency, they have several fundamental limitations. For example, Rychlik et al.'s model does not have a theoretical foundation for prediction of the optimal annealing temperature and their empirical correlation is purely based on a limited number of experiments. Stolovitzky and Cecchi's (7) and Velikanov and Kapral's (8) kinetic models did not account for the sequence dependence of amplification kinetics or were limited to a single step of the reaction.

A so-called state space model is required for dynamic optimization of DNA amplification. State space models are systems of differential equations that, when solved, describe the dynamics of the system, along with algebraic constraints and specified parameters (e.g., rate parameters such as activation energies and pre exponential factors) whose values are either predicted based on

first-principles theory, independently measured in offline experiments, or indirectly estimated through online measurement of observable quantities during the evolution of the system. Across the published literature, proposed state space models (10, 11, 12, 13, 14) give only poor estimates of the PCR amplification efficiencies. This is because no generalization has been made regarding the dependence of kinetic parameters on both i) the DNA sequence and ii) temperature. None of these kinetic models are both sequence- and temperature-dependent. It is quite evident from the nearest neighbor method, which can be used to calculate the DNA annealing reaction free energy, that the equilibrium constant of DNA hybridization is temperature and sequence dependent. Datta and Licata (15) reported temperature-dependent equilibrium dissociation constants for the enzyme binding reaction. Huang *et al* (16) and Innis *et al* (17) reported temperature-dependent enzyme extension reaction rates. Therefore, the kinetic parameters of the three steps of PCR are highly dependent on the sequence composition and temperature of the reaction. Accurate and computationally efficient sequence-dependent state space models, which are essential to solve such problems, require a combination of fundamental biophysical modeling with dynamical systems theory. This notion of sequence-dependent modeling of the kinetics of biochemical reaction networks, which has various applications in dynamical systems biology, is introduced here as one of the contributions of the present work.

In this paper, we develop the first sequence-dependent kinetic model for PCR reactions that is suitable for engineering control, validating this state space model through comparison to experimental data. First principles models are essential for proper prediction of DNA kinetic rate parameters for any arbitrary DNA sequence. The model introduced herein is based on quantitative biophysical modeling of DNA melting, annealing, and polymerization, which together enable a mapping of a given DNA sequence and polymerase enzyme onto temperature-dependent kinetic rate constants for the DNA amplification reaction. Such sequence-dependent modeling of amplification kinetics has been enabled, in part, by recent developments in the theory of DNA hybridization kinetics (18). One benefit of such complete state space models for PCR is the ability to achieve similar or enhanced amplification efficiencies and specificities in greatly reduced time, through the exploitation of dynamic processes – such as simultaneous annealing and extension - that are not represented in simplified models of DNA amplification, but which play a major role in determining the evolution of chemical species. Prospects for the

application of these sequence-dependent models in the formulation of optimal control problems that can enable the computation of optimal cycling strategies, for any specified objective, are discussed.

2. Kinetic Model for PCR

A kinetic model of PCR consists of kinetic models of melting, annealing, enzyme binding and extension reactions. In this work we have developed a sequence- and temperature dependent state space model for PCR and analyzed its kinetics.

2.1 Annealing Kinetics

Reaction R_1 represents an annealing reaction between the single strands (S) and primers (P).

$$S + P \underset{k_r}{\overset{k_f}{\leftrightarrow}} SP \qquad (R_1)$$

Marimuthu and Chakrabarti (18) developed a sequence and temperature dependent method to estimate the annealing reaction rate constants. Here we summarize their method that needs to be followed to estimate annealing rate constants.

2.1.1. Estimation of annealing rate constants.

1. Determine the overall Gibbs free energy and hence the equilibrium constant $K_{annealing}$ for a given sequence at the chosen annealing temperature using the Nearest Neighbor model.

$$K_{annealing} = \frac{k_f}{k_r} = \exp\left(-\frac{\Delta G_{annealing}}{RT}\right)$$

2. Determine the relaxation time, a characteristic time constant that determines the evolution of reaction coordinates toward equilibrium, at a chosen temperature using either one- or two-sided melting. Denoting by the probability of i bases of the primer being hybridized, for a homogeneous sequence the master equation is that of a biased one-dimensional random walk with partially reflecting boundary conditions:

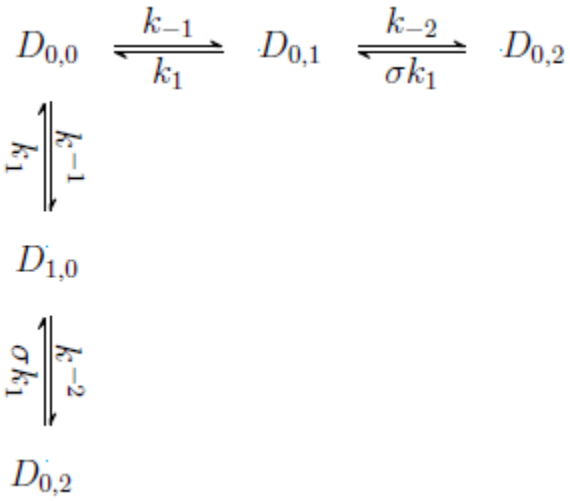
$$\frac{\partial}{\partial t}\wp(0,t) = -k_1\wp(0,t) + k_{-1}\wp(1,t)$$

$$\frac{\partial}{\partial t}\wp(i,t) = -(k_1 + k_{-1})\wp(i,t) + k_1\wp(i-1,t) + k_{-1}\wp(i+1,t), \quad i = 1,2,....n-1$$

$$\frac{\partial}{\partial t}\wp(n,t) = k_1\wp(n-1,t) - k_{-1}\wp(n,t)$$

We briefly review the method of Marimuthu and Chakrabarti (18) for calculating the relaxation time of such systems, including the more general case of two-sided heteropolymer melting.

- For the given primer length write the reaction mechanism; for example, for $N = 2$,



- Obtain the values of the rate constants and other parameters as explained in Marimuthu and Chakrabarti (18) and form the following state space matrix based on the above reaction mechanism

$$A = \begin{bmatrix} -\left(k^{-1} + k_{-1}\right) & k_1 & k_1 & 0 \\ k_{-1} & -\left(k_{-2} + k_1\right) & 0 & \sigma\left[D_{0,2eq}\right]k_1 \\ k^{-1} & 0 & -\left(k^{-2} + k_1\right) & \sigma\left[D_{0,2eq}\right]k_1 \\ 0 & k_{-2} & k^{-2} & -2\sigma\left[D_{0,2eq}\right]k_1 \end{bmatrix}$$

- Calculate the Eigenvalues λ_i of A.
- Calculate the relaxation time as per the following equation:

$$\tau = -\frac{1}{\max \lambda_i}$$

3. Relaxation time for the reaction R_1 in terms of k_f and k_r can be expressed as

$$\tau = \frac{1}{k_f \left(\left[S_{eq} \right] + \left[P_{eq} \right] \right) + k_r}$$

   $[S_{eq}]$ and $[P_{eq}]$ should be determined based on the initial concentrations of single strands and primers that are used to determine relaxation time in step 2.

4. Solve the two equations in steps 1 and 3 to determine k_f and k_r.

Fig. 1 shows the above procedure as a flowchart.

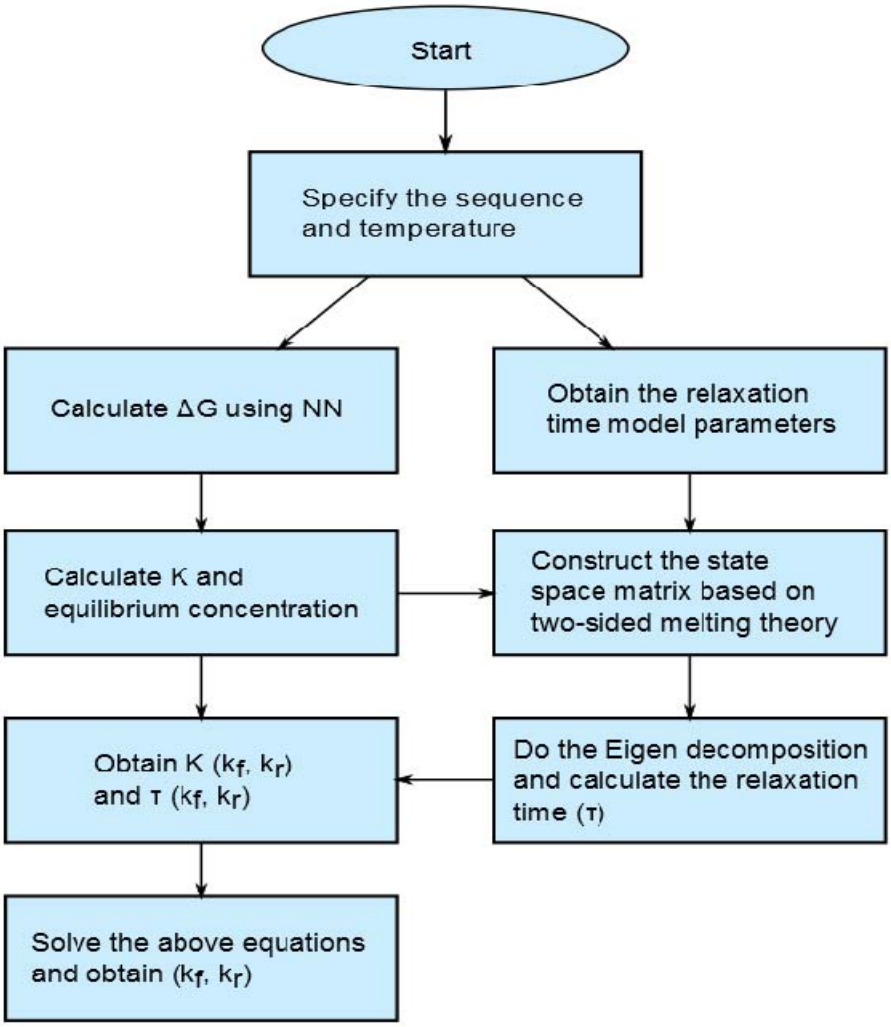


Figure 1: Flowchart for the estimation of sequence and temperature dependent annealing rate constants.

Using the above steps we have estimated the forward and reverse rate constants for a set of primers and the reaction parameters are given in Table 1. Arrhenius plots for the annealing rate constants k_f and k_r have been given in Fig. 2.

Sequence	$(E_a/R)_f$ (K^{-1})	$(E_a/R)_r$ (K^{-1})	$(k_0)_f$	$(k_0)_r$	T_m (^{0}C)
GCTAGCTGTAACTG	-7385	46341	2×10^{-4}	5×10^{63}	41
GTCTGCTGAAACTG	-8202	45926	2×10^{-5}	10^{63}	44

Table 1: Rate parameters of primers. Subscript *f* and *r* denotes forward and reverse rate constant, respectively.

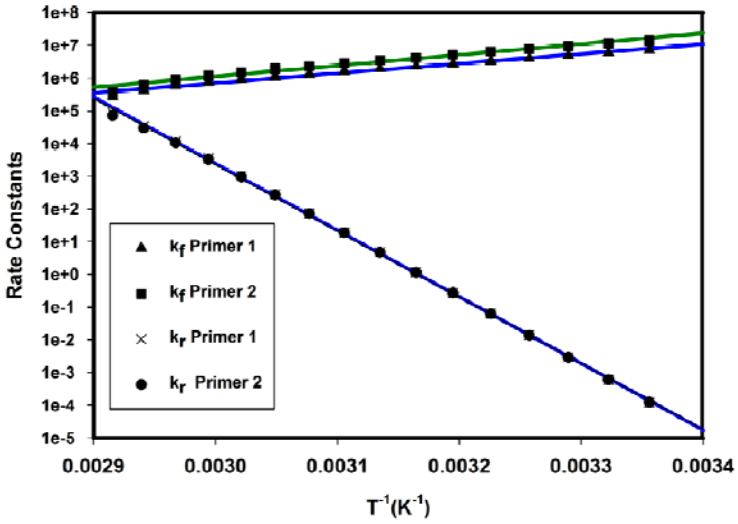


Figure 2. Arrhenius plot of the forward (k_f) and reverse (k_r) rate constants for the primer set 1.

From the melting curve (not shown) of the chosen primers it can be inferred that 100 % equilibrium conversion for reaction R_1 can be achieved at any temperature less than 32 ^{0}C. A very low annealing temperature could form mismatched products; therefore, the minimum annealing temperature is fixed to be 32 ^{0}C. This also reduces the range of the PCR operating temperatures, reducing the transition time between annealing and the other two steps of the PCR reaction.

2.2 DNA Melting

$$DNA \underset{k_{-m}}{\overset{k_m}{\leftrightarrow}} S_1 + S_2 \qquad (R_2)$$

DNA melting (R_2) is the reverse reaction of DNA hybridization, of which annealing reaction R_1 is an example. The kinetics of short DNA melting can be modeled using the 'all or none' or 'two state' model (19, 20). Based on this model, a DNA is assumed to be either in a single stranded or double stranded state; this assumption is valid only when the number of base pairs is less than 50 (21). Long DNA melting obeys co-operative melting, in which different regions of a DNA melt simultaneously in a different manner. The Poland-Scheraga (PS) model (22, 23, 24) can be used to predict this behavior and identify the different regions that can melt independently. However, to the best of our knowledge, the kinetics of long DNA melting has not been investigated. Mehra and Hu (10) assumed a rate constant that corresponds to the melting of a short DNA. Gevertz *et al* (11) and Stolovitzky and Cecchi (7) assumed that DNA melting is always 100% efficient and neglected the melting step in the overall PCR model. Unlike the annealing step, where primer annealing, enzyme binding and extension reactions can occur simultaneously, in the melting step , only the melting reaction can occur due to very high temperature. Moreover, as long as the given DNA does not form any secondary structures, it can melt completely. iIn our numerical studies, we have hence also assumed that the DNA melting reaction is always 100% efficient.. This approach allows us to simplify the treatment of the melting step; however, it does not consider the effect of template melting during the extension step at 72 ^{0}C which is moderately a high temperature at which a long DNA may melt. Furthermore, in applications like COLD PCR, the melting temperature is lower than the typical PCR DNA melting temperatures (25). In order to account for these factors, the temperature-dependent melting rate constants need to be estimated. Although we did not consider these effects in the numerical simulations of the present study, we present a method that can be used to model the kinetics of long DNA melting.

2.2.1. Statistical Mechanical Model for the Kinetics of the Melting of a long DNA

As mentioned above, long DNA molecules melts based on cooperative melting. Using the Poland-Scheraga (P-S) model, a given DNA sequence can be divided into 5 discrete domains: 1) Internal loops, 2) Pre-existing coils, 3) Expansion of loops, 4) Coalescence of neighboring loops,

and 5) Ends. There are many numerical methods and software such as MELTSIM (26, 27, 28) developed to identify the aforementioned domains and solve P-S model for a given long DNA sequence. Once these domains are found, for each domain an overall stability constant is calculated based on the following equation.

$$K_{loop} = \sigma_c f(N) \prod_{i=1}^{N} s_i \tag{1}$$

In the above Eq. (1), σ_c is referred as a cooperativity parameter (26) and it is different from the nucleation parameter σ that has been discussed in annealing model. σ_c is a penalty to the statistical weight for melting of the domain due to the free energy cost of dissociating an internal base pair. Both σ_c and $f(N)$ have been universally estimated and Blake *et al* (27) provided the expressions to calculate them. Each of the above regions melts independently based on the two-sided melting theory (29). These regions can be identified and the overall stability constant can be estimated as explained above. With this information in hand, the relaxation time of melting of the overall DNA can be found as follows.

2.2.2. Relaxation Time of a Long DNA

As shown in Section 2.1, an exact state space model for the melting of each base pair in each domain can be formulated. Since we know all the domains based on the P-S model, now the state space systems of each domain can be connected to find the state space system of overall DNA melting. The state space matrix of each domain will be coupled to only one other block, and in the following way: assuming that domain i+1 melts after domain i, only the fully molten domain i / fully hybridized domain i+1 state will be coupled to all the fully molten domain i /single-base dissociated domain i+1 states. Based on the type of domain, the state space system for each domain can be modeled using one-sided or two-sided melting theory (18). Once the state space matrix of the overall DNA melting is formulated, from the largest eigenvalue of the state space matrix, the relaxation time can be estimated. The following steps can hence be used to find the rate constants of long DNA melting.

- Using MELTSIM, identify different domains for a given DNA sequence.
- Order the domains based on their melting temperatures (ascending order).

- Construct the state space matrices for each domain based on two-sided or one-sided melting theory, as explained in the Supporting Information.
- Associate σ, the nucleation parameter, with individual states in the last block according to the method described in reference (18).
- Diagonalize each block and rank order the eigenvalues of all blocks.
- For each domain find the relaxation time using $-1/\max(\lambda_i)$.

Compare all the relaxation times; the maximum value of the relaxation time is the relaxation time of the whole DNA sequence. One may then compute the forward and reverse rate constants for long DNA melting using the relaxation time and the overall stability constant of the sequence, as described in Section 2.1.

2.3. Enzyme Binding and Extension Reaction Kinetics

In the extension step, enzymatic addition of nucleotides converts the duplex primer-template complex (SP) into a complete dsDNA. Both deterministic as well as stochastic (8) approaches have been proposed to develop a model for the extension reaction. Velikanov and Karpal (8) presented the following chemical master equation (CME), a probabilistic description of the extension reaction system, together with its analytical solution:

$$\frac{\partial}{\partial t}P(l,t) = w_{l-1\to l}P(l-1,t) - w_{l\to l+1}P(l,t) \tag{2}$$

where $P(l,t)$ denotes the probability distribution of the duplexes with l base pairs added through the extension reaction, and $w_{l-1\to l}$ denotes the transition probability rate of nucleotide addition to the l-1 base pair duplex.

Although the solution of the above CME can provide the time required to complete an extension reaction, this formulation is not useful in the present context because

- It omits enzyme dissociation/processivity as one of its major drawbacks; hence it only applies to perfectly processive polymerases (see below for discussion of processivity). Since thermostable enzymes are not perfectly processive, it cannot be applied to PCR.

- It cannot be integrated with the models for the other steps of PCR. Therefore, it is impossible to analyze annealing, enzyme binding and extension reactions simultaneously.

Therefore, we consider an alternative approach with an appropriate reaction mechanism to develop a model for the enzyme binding and extension reactions.

2.3.1. Reaction Mechanism

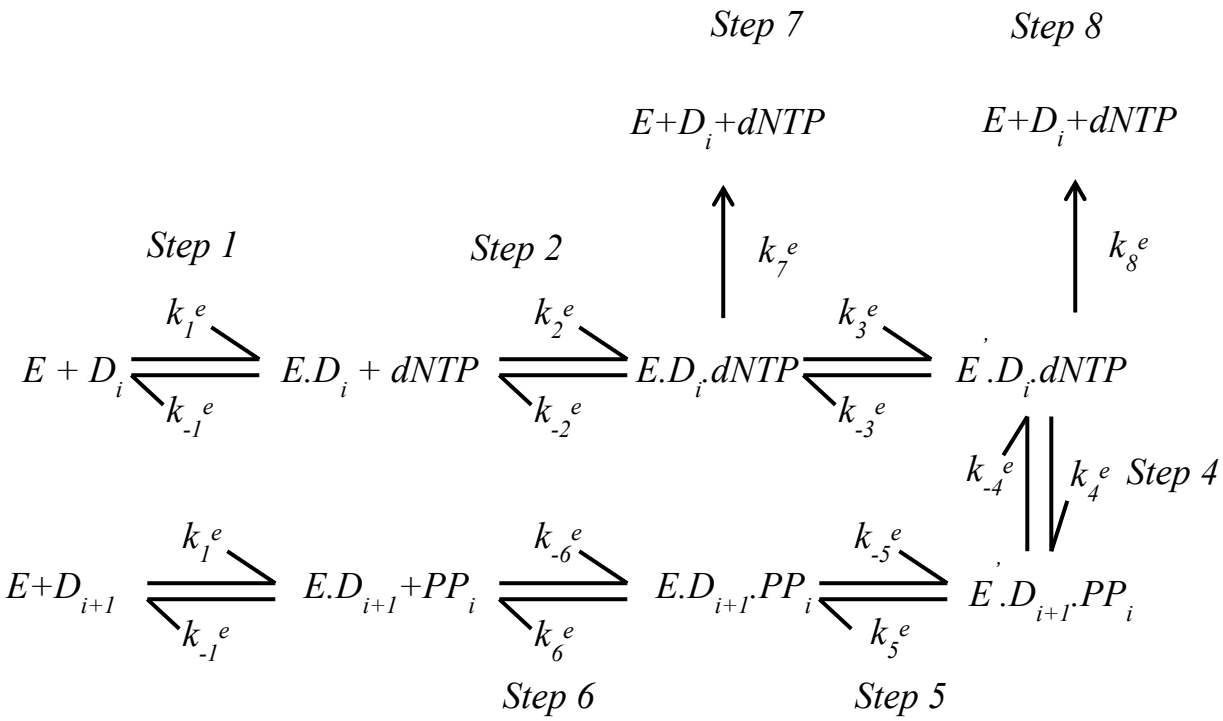


Figure 3. A general reaction mechanism of Enzymatic Primer Extension reaction. (This figure has been reproduced from Brown et al. (35).)

There are several reaction mechanisms proposed for the enzyme binding and extension reactions (31, 32, 33, 34) and Fig. 3 represents a general reaction mechanism (35). In step 1, enzyme binds with D_i molecule to form a binary complex $E.D_i$. In step 2, a deoxynucleoside triphosphate (*dNTP*) binds with $E.D_i$ to form a ternary complex, *E.Di.dNTP*, which undergoes a protein conformational change in step 3 and forms $E'.D_i.dNTP$. In step 4, the nucleotide is incorporated and a pyrophosphate molecule is released from D_i, and as a result $E'.Di_{+1}.PP_i$ is formed.

$E'.D_{i+1}.PP_i$ undergoes a conformational change in step 5, leading to the formation of $E.D_{i+1}.PP_i$. In step 6, PP_i is completely released from $E.D_{i+1}.PP_i$ and $E.D_{i+1}$ is formed. Finally the dissociation of $E.D_{i+1}$ produces D_{i+1} and E. Besides these steps, there are parallel dissociation reactions represented by step 7 and 8 that may also occur.

Kuchta *et al* (30), Patel *et al* (31), Brown and Suo (35), Capson *et al* (36), and Fiala *et al* (37) studied the extension reaction kinetics for DNA polymerase I *Klenow*, T7 DNA polymerase, *S. solfataricus* P2 DNA polymerase B1, T4 gene 43 protein, and *S. solfataricus* P2 DNA polymerase IV, respectively, at either 20 ^{0}C or 37 ^{0}C. Using their rate constant data, we simplify the above reaction mechanism. Step 6 is the last step of the reaction mechanism that produces $E.D_{i+1}$. According to Patel *et al* (31), k_6^e = 1000 s^{-1} and k_{-6}^e = 0.5 $(\mu M)^{-1}$ s^{-1}. These rate constant values suggest that the association of $E.D_{i+1}$ with PP_i is essentially impossible given the PPi concentrations in typical DNA amplification reactions. In addition to this, comparing k_{-6} with k_1 (11 $(\mu M)^{-1}$ s^{-1}) and k_2^e (> 50 $(\mu M)^{-1}$ s^{-1}), it can be considered negligible. Hence, the final step 6 is irreversible with a rate constant k_6^e. Step 3, 4 and 5, which are all a first order reversible reactions, represent the conformational change of a ternary complex. Their rate constants values are higher than k_2^e (31), and the forward rate constants for each step are higher than the corresponding reverse rate constants (31). Hence, the overall dynamics are controlled by step 2, which forms a ternary complex, $E.D_i$.dNTP, and the final step is irreversible. Thus, as proposed by Boosalis *et al* (32), Mendelman *et al* (33) and Huang *et al* (34), the above reaction mechanism can be represented using the simplified reaction schemes given by reaction scheme R_3.

$$E + D_i \underset{k_{-1}^e}{\overset{k_1^e}{\rightleftharpoons}} E.D_i + dNTP \underset{k_{-2}^e}{\overset{k_2^e}{\rightleftharpoons}} E.D_i.dNTP \xrightarrow{k_{cat}} E.DNA_{i+1} + PP_i \underset{k_{-1}^e}{\overset{k_1^e}{\rightleftharpoons}} E + D_{i+1} \qquad (R_3)$$

We now show how the kinetic parameters k_1^e, k_{-1}^e, k_{cat} and $K_N = (k_{-2}^e + k_{cat})/k_2^e$ in this reaction scheme can be estimated for any polymerase using polymerase processivity and initial rate experiments.

2.3.2. Single-hit conditions:

Under so-called single hit conditions, enzyme concentrations are sufficiently low that the probability of re-association is approximately zero. Therefore, they do not allow enzyme re-

association. Hence enzyme-template association occurs only during the initial equilibration of enzyme with SP. Thus, the following reaction scheme for the addition of *n* base pairs is written:

$$E + D_0(SP) \underset{k_{-1}^e}{\overset{k_1^e}{\rightleftharpoons}} E.D_0 + N \underset{k_{-2}^e}{\overset{k_2^e}{\rightleftharpoons}} E.D_0.N \xrightarrow{k_{cat}} E.D_1$$

$$E + D_i \xleftarrow{k_{-1}^e} E.D_i + N \underset{k_{-2}^e}{\overset{k_2^e}{\rightleftharpoons}} E.D_i.N \xrightarrow{k_{cat}} E.D_{i+1} \xrightarrow{k_{-1}^e} E + D_{i+1} \qquad i = 1, 2....n\text{-}1 \quad (\boldsymbol{R_4})$$

$$E + D_{n-1} \xleftarrow{k_{-1}^e} E.D_{n-1} + N \underset{k_{-2}^e}{\overset{k_2^e}{\rightleftharpoons}} E.D_{n-1}.N \xrightarrow{k_{cat}} E.D_n \xrightarrow{k_{cat}^{'}} E + D_n$$

Single hit conditions are used to estimate polymerase processivity parameters. . Processivity is defined as the expected number of nucleotides incorporated per DNA-enzyme binding event, and is discussed further below.

2.3.3. Processivity of an enzyme

Let *i* index the sequence positions on the template. In a Markov chain formulation of dissociation, the index *i* at which dissociation occurs is called the stopping index and is denoted i_{off}. Let *p* denote the conditional probability of the polymerase *not* dissociating at position *i,* given that it was bound to the template at position/time *i*-1. The probability of dissociation at position *i* is then

$$p_{off}(i) = (1-p)\,p^{i-1} \tag{3}$$

p is called the microscopic processivity parameter. The expected position of dissociation of the polymerase (expected stopping index), called the processivity, can be written as

$$E[i_{off}] = \frac{1}{1-p}$$

$E[i_{off}]$ is typically reported as the processivity instead of the microscopic processivity parameter. The above expression is derived for a template of infinite length. Usually, in processivity experiments long templates are used to estimate *p*. For finite length,

$$p_{off}(i) = 1 - \sum_{i=0}^{n-1} (1-p)\,p^{i-1} .$$

For heterogeneous templates, *p* will vary with position. From processivity experiments, one can obtain the *p* at each position since we will have

$$p_{off}(i) = (1 - p_i)\prod_{j=0}^{i-1} p_j \quad \forall i$$

These equations can be used to solve uniquely for each p_i. However, it is impractical to do processivity experiments for each new template. Hence, one can do processivity experiments on templates with different types of nearest neighbor motifs (including hairpins) for a given polymerase, and then these nearest neighbor processivity parameters can be used in modeling of the processivity for an arbitrary sequence.

2.3.4. Relationship between Processivity and Enzyme binding/Extension rate constants.

Now, at a fixed temperature, we seek a relationship between processivity of an enzyme and the rate constants of the reaction scheme ***R4***. In order to do this, we write the state space model for the reaction scheme ***R4***. We omit $E.D_n \xrightarrow{k'_{cat}} E + D_n$ from the state space model for simplicity as it does not affect equilibrium and we are not estimating the corresponding rate constant. Since the substrate, dNTP, is always in excess compared to enzyme, Michaelis-Menten (MM) kinetics is valid and hence the steady state assumption for the intermediate concentration is valid. Therefore,

$$\frac{d}{dt}[E.D_i.N] = -[E.D_i.N](k_{-n} + k_{cat}) + k_n[E.D_i][N] = 0 \Rightarrow [E.D_i.N] = \frac{[E.D_i][N]}{K_N} \qquad (6)$$

where $K_N = \frac{k_{-n} + k_{cat}}{k_n}$. Let $k = \frac{k_{cat}}{K_N}[N]$; then the state space matrix A of the reaction scheme ***R4*** is given as

$$\frac{dx}{dt} = Ax \Rightarrow \frac{dx}{d\left(t\left(k+k_{-1}^{e}\right)\right)} = \frac{1}{k+k_{-1}^{e}} A \Rightarrow \frac{dx}{dt'} = A'x\,,\quad x = [E.D_0, D_0, E.D_1, D_1, ..., D_{n-1}, E.D_n]^T \quad (7)$$

$$A' = \begin{bmatrix} -1 & 0 & 0 & 0 & 0 & & 0 \\ \frac{k_{-1}^{e}}{k+k_{-1}^{e}} & 0 & 0 & 0 & 0 & & 0 \\ \frac{k}{k+k_{-1}^{e}} & 0 & -1 & 0 & 0 & & 0 \\ 0 & 0 & \frac{k_{-1}^{e}}{k+k_{-1}^{e}} & 0 & 0 & \cdots & 0 \\ 0 & 0 & \frac{k}{k+k_{-1}^{e}} & 0 & -1 & & 0 \\ & & \vdots & & & \ddots & \\ 0 & 0 & 0 & 0 & 0 & & 0 \end{bmatrix};$$

where

$k_{-1}^{e}dt$ = conditional probability of transition from state $E.D_i \xrightarrow{k_{-1}^{e}} E + D_i$ in time dt

$k\,dt$ = conditional probability of transition from state $E.D_i \rightarrow E.D_{i+1}$ in time dt.

In a single molecule continuous time Markov chain formulation, Eq. (7) can be written in terms of the probability distribution of states $\left[\wp_{D_0}, \wp_{E.D_0}, ...\right]^T$ instead of the vector of species concentrations. An equivalent master equation formulation is:

$$\frac{\partial}{\partial t}\wp(0,t) = -(k+k_{-1}^{e})\wp(0,0); \frac{\partial}{\partial t}\wp(1,t) = k_{-1}^{e}\wp(0,0); \frac{\partial}{\partial t}\wp(2,t) = k\,\wp(0,0), ... \quad ; \wp(0,0) = 1$$

where $\wp(i,t)$ denotes the probability of the polymerase being in state i at time t.

The equilibrium distribution of this master equation can be obtained by solving for the generalized eigenvectors of the state space system Eq. (7). It is found that this distribution has the form specified by Eq.(3) with $p_{off}(i) \equiv p_{off}(i, t=\infty)$ and the following value of the microscopic processivity parameter:

$$p = \frac{\frac{k_{cat}}{K_N}[N]}{\frac{k_{cat}}{K_N}[N] + k_{-1}^{e}}$$

Hence

$$\frac{k_{-1}^{e}}{\frac{k_{cat}}{K_N}[N]+k_{-1}^{e}} = 1-p \Rightarrow k_{-1}^{e} = \frac{\frac{k_{cat}}{K_N}[N](1-p)}{p} \tag{8}$$

and Eq. (8) can be written as follows in terms of processivity:

$$k_{-1}^{e} = \frac{k_{cat}}{K_N}[N]\left(\frac{1}{E\left[i_{off}\right]-1}\right) \tag{9}$$

As per Eq. (9), if $\frac{k_{cat}}{K_N}$ and processivity of a polymerase are known at a specific temperature, it is possible to estimate k_{-1}^{e} for any polymerase. Eq. (9) is valid under the approximations applied in the derivation of (R3). Moorthy *et al* (S.Moorthy, K.Marimuthu, and R. Chakrabarti, unpublished data) carry out a comprehensive analysis of these approximations and consider more general models. For each such model, an equation for k_{-1} analogous to (R_3) can be derived based on the associated single hit *A* matrix, in terms of processivity and other model parameters.

Moorthy *et al* (S.Moorthy, K.Marimuthu, and R. Chakrabarti, unpublished data) estimated $\frac{k_{cat}}{K_N}$ for *Taq* polymerase at various temperatures based on a bireactants MM kinetics formulation. Fig. 5c shows the temperature dependent extension rate constant $\frac{k_{cat}}{K_N}$. In Eq. (9) k_{-1}^{e} and $\frac{k_{cat}}{K_N}$ are concentration independent terms and hence, the processivity $E\left[i_{off}\right]$ or the conditional probability *p* depends on [*N*]. In order to use Eq. (9) to estimate k_{-1}^{e}, [*N*] and $E\left[i_{off}\right]$ should be consistent or one should use the value [*N*] at which $E\left[i_{off}\right]$ is estimated. Wang *et al* (38) and Davidson *et al* (39) determined the value of $E\left[i_{off}\right]$ at a specific temperature and nucleotide

concentration. The following Table provides the values of $E\left[i_{off}\right]$ and the conditions at which they were measured.

Reference	Temperature (^{0}C)	Nucleotide concentration [N] (μM)	$E\left[i_{off}\right]$
Wang *et al* (36)	72	250	22
Davidson *et al* (37)	60	800	50-80

Table 2: Processivity of Taq polymerase

Using the above values, k_{cat}/K_N and Eq. (9), we have estimated k_{-1}^{e} at 60 ^{0}C and 72 ^{0}C, respectively. We have the k_{-1}^{e} for *S. solfataricus* P2 DNA polymerase B1 (37) at 37 ^{0}C and we use the same value for *Taq* polymerase enzyme as their equilibrium constants are of the same order of magnitude. Thus, we can obtain estimates of k_{-1}^{e} at three different temperatures and using these, an Arrhenius relationship is fitted as shown in Fig. 5b to estimate the temperature dependent dissociation rate constant k_{-1}^{e}. Fig. 4 explains the steps involved in enzyme binding and extension model parameter estimation.

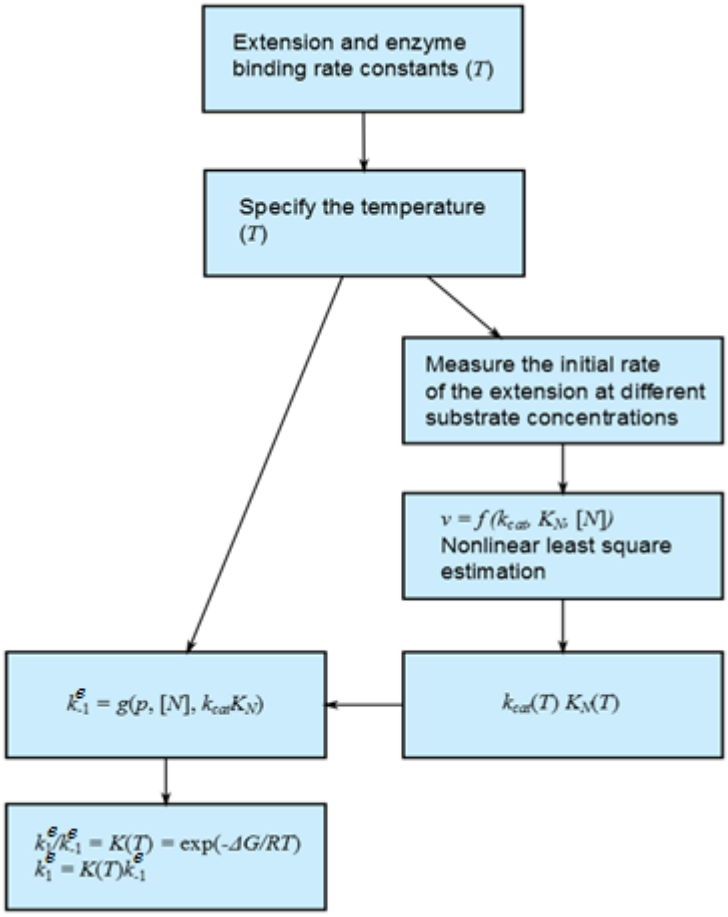
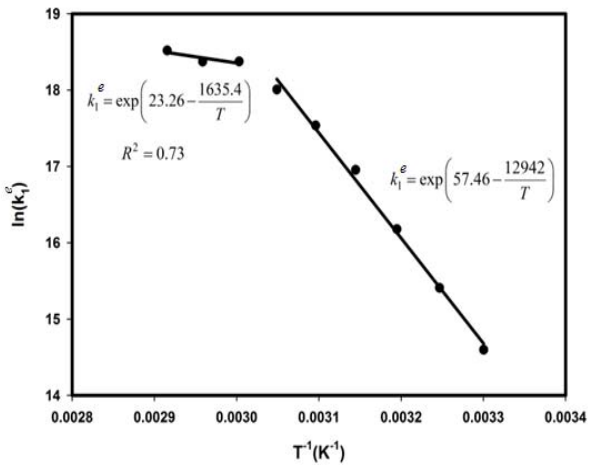
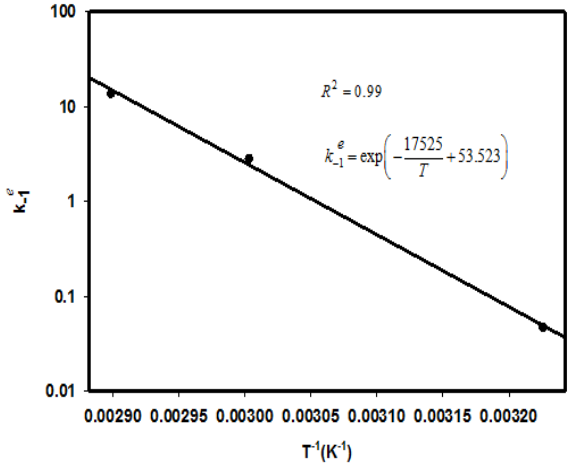


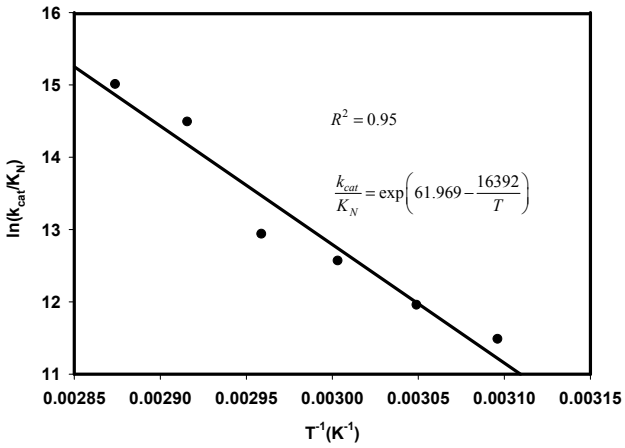
Figure 4: Estimation of enzyme binding and extension rate constants.



(a)



(b)



(c)

Figure 5: Temperature dependence of a) Enzyme dissociation rate constant; b) Enzyme binding rate constant; c) Extension reaction rate constant for *Taq* polymerase

Equilibrium thermodynamic analysis for the enzyme binding reaction has been done extensively with *Thermus aquaticus* enzyme by Datta and LiCata (15). They estimated the temperature dependent equilibrium constant which is the ratio k_1^e/k_{-1}^e. Therefore, based on the temperature dependence of k_{-1}^e and $K_{binding}$, k_1^e has been estimated and its Arrhenius plot is shown in Fig. 5a.

3. Analysis of PCR Kinetics

Using the kinetic model developed in Section 2, we seek an optimal temperature vs. time profile for DNA amplification. In order to do this, we first analyze the kinetic model in this Section to assess the importance of such simulations in making accurate predictions of the amplification efficiency of PCR reactions. We note that the conventional picture of PCR kinetics assumes a single reaction is rate-limiting for each step estimates reaction temperatures and times for each step based on this assumption without solving the associated state equations.

Except for the annealing temperature, the reaction conditions are the same as the typical PCR conditions recommended, for example, by Invitrogen (40). Reactions R_1, R_2, R_3 and R_4 have been written for a simplex PCR reaction and they are given along with their state equations in Appendix A1 and A2. In Figure A.1, we have shown which rate constants are sequence and temperature dependent. We summarize the simulation results as follows (data not shown here).

- If the ratio of single strand concentration to primer concentration (SP ratio) S_0/P_0 is <1, then the annealing reaction is nearly instantaneous.
- If the SP ratio is close to 1, then there is a transient behavior in the evolution of S_1P_1.

When the SP ratio is low, since the primer concentration is very high compared to the single strand concentration, the primer molecule easily binds to the single strand molecule and does not allow single strands to anneal to each other. On the other hand, when the SP ratio is 1, since the primer and single strand molecules are equal in concentration, there is a competition between them to anneal to their respective complementary sequences. Since the single strands participate in a two-way competition with both primers, they eventually lose in this competition. Thus, the annealing reaction is not the rate-limiting step during the early stages of PCR but it may become the rate limiting step towards the end of PCR.

In PCR, study of the annealing reaction separately may be misleading in deriving conclusions about the optimal annealing time. From Section 2.3, it is evident that enzyme binding can occur at annealing temperatures. This can affect the annealing and hence the overall dynamics of the PCR. Datta and LiCata's (15) experiments reveal that the Gibbs free energy of the enzyme

binding reaction has its minimum around 50 ^{0}C. In a conventional PCR model, however, enzyme binding is considered to occur during the extension reaction (10, 11). In the next section, we motivate fully time-varying state space models by showing that simultaneous annealing and extension reactions result in significant differences in reaction efficiency that can be exploited through such modeling. These effects are not captured in conventional models of PCR kinetics.

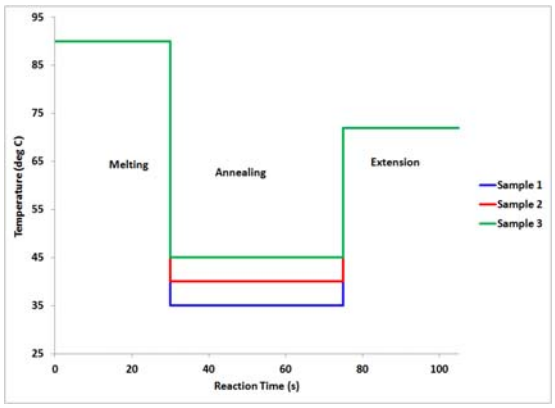
3.1. Combined Annealing and Extension

At any instant, since all the reactants for annealing, enzyme binding and extension reactions are available in the reaction mixture, these reactions can in principle occur simultaneously. The combined annealing and extension model allows us to simulate arbitrary PCR reaction cycling protocols that do not follow the standard 3-step scheme, hence extending beyond the types of "on-off" behavior commonly assumed in models of PCR. Due to this phenomenon, molecular biologists often run PCR reaction with only two steps per cycle - one step for melting and one step for annealing/extension. However, there is no quantitative prescription available for the temperatures of the annealing/extension steps. In order to provide such prescriptions, in this section we do not distinguish between the annealing and extension steps and solve the kinetic equations corresponding to all these steps together for the overall reaction time. This is one of the main reasons for the need for temperature-dependent rate parameters. The reaction conditions are the same as those in a typical PCR. Annealing and extension times are fixed to be 45 and 30 seconds, respectively, and the length of the target DNA is assumed to be 1000 base pairs (bp). The extension reaction temperature is 72 ^{0}C. At a given time, since one of the 3 steps of PCR is kinetically dominating, the rates of the all the reactions are not uniform. This difference in reaction rates creates a stiff state space system that needs to be solved carefully. We used the MATLAB routine *ode15s* to solve this system of stiff differential equations.

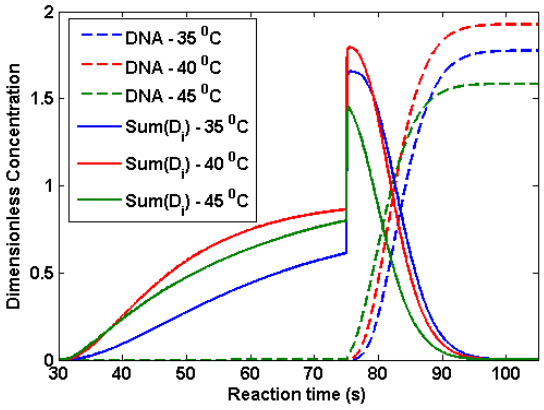
Although the annealing reaction is very fast at low temperatures, its efficiency is determined by the kinetics of the enzyme binding reaction. Therefore, even at high annealing temperatures at which the equilibrium conversion of the annealing reaction in the absence of enzyme binding is low, it is possible to obtain 100% overall efficiency. The evolution of single strands, single strand-primer duplex and final DNA for a single cycle at the annealing temperatures 35 ^{0}C, 40 ^{0}C and 45 ^{0}C are presented in Fig 6b. The temperature cycling profile for each annealing temperature is shown in Fig 6a. At an annealing temperature of 35 ^{0}C the annealing reaction can

reach 100 % equilibrium conversion. The equilibrium conversion of the enzyme binding reaction at this temperature is less than 70%. Nevertheless, the overall PCR conversion is more than 70% at 35 ^{0}C. This is due to the combined annealing, enzyme binding and extension reactions. The extension reaction (or subsequent shifts the equilibrium of the enzyme binding reaction, allowing more enzymes bind to the SP duplexes. The extension reaction rate increases when temperature increases.

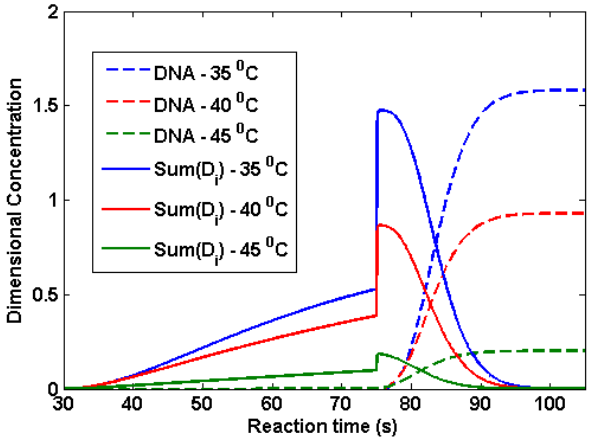
In the annealing step, all SP is converted into $E.D_i$ molecules, which can then dissociate into D_i. During this step, since the enzyme dissociation rate constant is comparable to the extension rate constants, $E.D_i$ molecules dissociate. Fig. 6b shows the profile of the sum of concentrations of D_i. When the temperature of the reaction is increased to 72 ^{0}C during the extension step, the equilibrium of $E.D_i$ dissociation is disturbed by the rapid nucleotide addition and eventually all D_i molecules are converted to $E.D_i$, which are in turn converted into target DNA. The melting temperatures of the primers in this study (Table 1) are less than 35 ^{0}C, and as a result of this, 45 ^{0}C annealing temperature did not produce more DNA. Since the rate constants of the enzyme binding and extension reactions at 35 and 40 ^{0}C are comparable, the the evolution of the DNA molecules at these two temperatures is similar. However, at 40 ^{0}C, the overall reaction is faster.



(a)



(b)



(c)

Figure 6. a) Three different temperature cycling samples. b) Transient behavior of reaction constituents (D_i and DNA molecules) for Primer set 1. c) Transient behavior of reaction constituents for Primer set 2. In both cases the annealing temperatures are 35 ^{0}C, 40 ^{0}C and 45 ^{0}C and the length of the target is 1000 bp. Annealing and extension times are 45 and 50 seconds, respectively. Primer, enzyme, and dNTP concentrations are 0.2 µM, 10 nM, and 800 µM, respectively. Since melting was assumed to be 100% efficient and melting dynamics were not simulated in these studies, species concentrations in b,c) are plotted starting with the annealing step.

We have repeated the above analysis for a different set of primer sequences (primer set 2) of the same length and the same reaction conditions. The sequences for primer set 2 are

Primer 1 ='AATAGCTGTAACTG';
Primer 2 ='TTCTTCTGAAACTG';

and the rate parameters have been calculated as explained in Section 2. Fig. 6c shows the evolution of DNA and the sum of D_i concentrations. Unlike primer set 1 (Table 1), in this case the favorable annealing temperature is 35 ^{0}C. Furthermore, for the same overall reaction time, the overall conversion is different for the two primer sets. This demonstrates that the sequence dependent kinetic model is important.

It should be noted that aforementioned results are applicable to the first cycle of a PCR reaction. In the above study, the enzyme concentration is in excess compared to the single strand concentration. If this condition does not hold, which is the case for the later stages of PCR, the kinetics could be very different. Also, during every PCR cycle, the target DNA concentration increases. As a result, the overall number of required nucleotide additions will also increase. Therefore, the reaction conditions that have been maintained during the initial stages of PCR may not be appropriate for the later stages of the PCR. This effect is more pronounced for longer sequences. Furthermore, note that even though our model considers the melting of SP molecules during the extension reaction, it does not consider the melting of D_i molecules during the extension reaction for the following reasons:

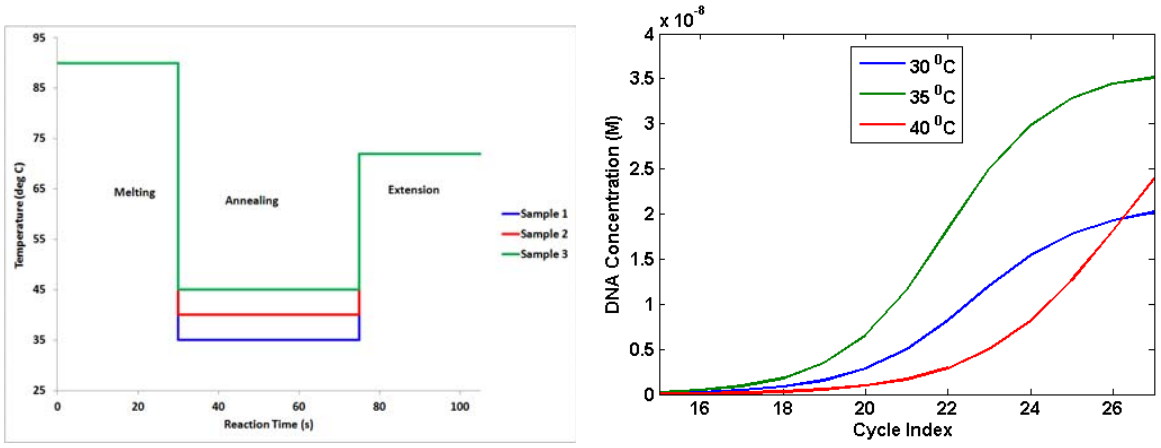
- Our simulation results suggest that even though $\sum_i [D_i]$ is considerable at the end of annealing step, as shown in Fig. 6b and 6c, the summation of the concentrations of D_1 to D_{15} molecules is negligible. Therefore, we assume that these molecules are not present in the reaction mixture.
- The stability/ melting temperature of a duplex increases when the number of base pairs increases. In the present study we have considered primers of length 14 base pairs. Therefore, since Di with i<15 can be neglected, the minimum number of base pairs in D_i is greater than 30 and we assume that these duplexes are stable at 72 ^{0}C.

Thus, in this Section we have established that

- Even during the annealing step, enzyme binding and extension reactions can occur simultaneously. Hence the state equations of annealing, enzyme binding and extension reaction should be solved together. As will be shown below, it is possible to exploit these simultaneous reactions to improve PCR reaction efficiency through appropriate model-implied choices of temperature cycling strategies.
- As an example, the kinetic model for annealing and extension can provide a quantitative prescription for two-step PCR (melting and combined annealing/extension).
- There should be an optimal annealing temperature at which the reaction is fastest and reaches 100% completion. Importantly, this temperature cannot be computed based on primer melting temperatures alone.

  When the length of the target DNA increases, and hence more nucleotides must be added, the annealing reaction temperature should be higher and reaction time should be increased. Again, the kinetic model for annealing and extension can provide a quantitative prescription for the optimal annealing time and temperature.

3.2. Geometric Growth of DNA



(a) **(b)**

Figure 7: a) Temperature profile for the first cycle at three different annealing temperatures. The same temperature profile is followed for all other cycles. b) Geometric Growth of DNA. Annealing and Extension reaction time in each step of the PCR are 45 and 30 seconds, respectively. The extension temperature is fixed to be 72 ^{0}C. Initial concentration of template, primers (primer set 1), enzyme, and nucleotide is 10^{-14} M, 0.2 µM, 10 nM and 800 µM, respectively.

For fixed annealing and extension times of 45 and 30 seconds, respectively, the state equations of the PCR reaction scheme were next solved at three different annealing temperatures, over multiple PCR cycles. The temperature cycling profile is shown in Fig. 7a. Fig. 7b shows the geometric growth of DNA concentration. When the annealing temperature is 35 ^{0}C, the DNA concentration saturates at 20 nM after 26 cycles, whereas at 40 ^{0}C annealing temperature, the DNA concentration is approximately equal to 35 nM after 27 cycles. Although the efficiencies at 35 and 40 ^{0}C annealing temperature are approximately same in the first cycle, as shown in Fig. 6b, when the cycle number increases, the final DNA concentration differs. At 45 ^{0}C, the final DNA concentration is higher than that at 35 ^{0}C. It should be noted that when the target DNA concentration is comparable with enzyme concentration, the dynamics of the PCR reaction depends on the annealing temperature. Ideally, the maximum concentration of the target DNA should be equal to primer concentration. Therefore, during the initial stage of the PCR, target DNA concentration is the limiting reactant. Once the target DNA concentration exceeds the

enzyme concentration, the latter is the limiting reactant. From Fig. 7b it is clear that in the second stage the PCR efficiency is lower and a different reaction condition needs to be maintained to improve the efficiency.

4. Conclusion

In this work, we have developed the first sequence- and temperature-dependent kinetic model for DNA amplification, through biophysical modeling of coupled DNA melting and polymerization processes. Using this model, the kinetics of PCR have been analyzed for various temperature cycling strategies. Based on the results of this kinetic analysis, the need for systematic optimization of temperature cycling strategies has been established. The theory of optimal control of dynamical systems (5) provides a framework for the computation of the optimal temperature cycling protocols for DNA amplification. Use of the proposed sequence-dependent kinetic model in a control-theoretic framework should enable determination of the optimal dynamic operating conditions of DNA amplification reactions, for any specified amplification objective. Through the application of this kinetic state space model, it may thus be possible to i) improve the overall amplification efficiency of the reaction by orders of magnitude for the same number of cycles; ii) substantially reduce the overall time of the reaction compared to conventional PCR protocols. Future work will consider an optimal control framework and solution strategy for maximization of the amplification efficiency, as well as control problems pertaining to other DNA amplification objectives. These include problems involving the co-amplification of multiple DNA sequences and the automated design of new types of PCR reactions. Models for such problems can be built on principles directly analogous to those presented in this paper.

Appendix

A.1 PCR Reactions

A.1.1. Melting

$$DNA \underset{k_{-m}}{\overset{k_m}{\rightleftharpoons}} S_1 + S_2$$

A.1.2. Annealing

$$S_1 + P_1 \underset{k_r^1}{\overset{k_f^1}{\rightleftharpoons}} S_1P_1$$

$$S_2 + P_2 \underset{k_r^2}{\overset{k_f^2}{\rightleftharpoons}} S_2P_2$$

$$S_1 + S_2 \underset{k_r^*}{\overset{k_f^*}{\rightleftharpoons}} DNA$$

A.1.3. Enzyme Binding and Extension

$$S_1P_1 \underset{k_{-1}^e}{\overset{k_1^e}{\rightleftharpoons}} E.S_1P_1 + N \underset{k_{-2}^e}{\overset{k_2^e}{\rightleftharpoons}} E.S_1P_1.N \xrightarrow{k_{cat}} E.D_1^1$$

$$S_2P_2 \underset{k_{-1}^e}{\overset{k_1^e}{\rightleftharpoons}} E.S_2P_2 + N \underset{k_{-2}^e}{\overset{k_2^e}{\rightleftharpoons}} E.S_2P_2.N \xrightarrow{k_{cat}} E.D_1^2$$

$$E + D_i \xleftarrow{k_{-1}^e} E.D_i + N \underset{k_{-2}^e}{\overset{k_2^e}{\rightleftharpoons}} E.D_i^j.N \xrightarrow{k_{cat}} E.D_{i+1}^j \xrightarrow{k_{-1}^e} E + D_{i+1}^j \qquad i = 1,2....n\text{-}1,\ j = 1\ \&\ 2$$

$$E + D_{n-1} \xleftarrow{k_{-1}^e} E.D_{n-1}^j + N \underset{k_{-2}^e}{\overset{k_2^e}{\rightleftharpoons}} E.D_{n-1}^j.N \xrightarrow{k_{cat}} E.D_n^j \xrightarrow{k_{cat}^{'}} E + D_n$$

A.2 Differential Equations for the PCR modeling.

A.2.1. Melting

$$\frac{d}{dt}[DNA] = k_{-m}[S_1][S_2] - k_m[DNA]$$

A.2.2. Annealing

$$\frac{d}{dt}[S_1P_1] = k_f^1[S_1][P_1] - k_r^1[S_1P_1] - k_1^e[E][S_1P_1] + k_{-1}^e[E.S_1P_1]$$

$$\frac{d}{dt}[P_1] = -k_f^1[S_1][P_1] + k_r^1[S_1P_1]$$

$$\frac{d}{dt}[S_2P_2] = k_f^1[S_2][P_2] - k_r^1[S_2P_2] - k_1^e[E][S_2P_2] + k_{-1}^e[E.S_2P_2]$$

$$\frac{d}{dt}[P_2] = -k_f^2[S_2][P_2] + k_r^2[S_2P_2]$$

$$\frac{d}{dt}[S_1] = -k_f^1[S_1][P_1] + k_r^1[S_1P_1] - k_{-m}[S_1][S_2] + k_m[DNA]$$

$$\frac{d}{dt}[S_2] = -k_f^2[S_2][P_2] + k_r^2[S_2P_2] - k_{-m}[S_1][S_2] + k_m[DNA]$$

$$\frac{d}{dt}[DNA] = k_{-m}[S_1][S_2] - k_m[DNA]$$

A.2.3. Enzyme Binding and Extension

$$\frac{d}{dt}[E.S_1P_1] = k_1^e[E][S_1P_1] - k_{-1}^e[E.S_1P_1] - k_{cat}\frac{[E.S_1P_1][N]}{K_N}$$

$$\frac{d}{dt}[E.S_2P_2] = k_1^e[E][S_2P_2] - k_{-1}^e[E.S_2P_2] - k_{cat}\frac{[E.S_2P_2][N]}{K_N}$$

$$\frac{d}{dt}[E.T_i] = k_1^e[E][T_i] - k_{-1}^e[E.T_i] + k_{cat}\frac{\{[E.T_{i-1}] - [E.T_i]\}[N]}{K_N}$$

$$\frac{d}{dt}[T_i] = -k_1^e[E][T_i] + k_{-1}^e[E.T_i]$$

$$T_i = D_1^1, D_1^2, D_2^1, D_2^2, \ldots\ldots\ldots\ldots D_{n-1}^1, D_{n-1}^2$$

$$\frac{d}{dt}\left[E.D_n^1\right] = \frac{k_{cat}}{K_N}[N]\left[E.D_{n-1}^1\right] - k_{cat}^{'}\left[E.D_n^1\right]$$

$$\frac{d}{dt}\left[E.D_n^2\right] = \frac{k_{cat}}{K_N}[N]\left[E.D_{n-1}^2\right] - k_{cat}^{'}\left[E.D_n^2\right]$$

$$\frac{d}{dt}[DNA] = k_{cat}^{'}\left\{[E.D_n^1] + [E.D_n^2]\right\}$$

$$\frac{d}{dt}[N] = -\frac{k_{cat}}{K_N}[N]\left\{\sum_{T=S_1P_1}^{D_{n-1}^1}[E.T] + \sum_{T=S_2P_2}^{D_{n-1}^2}[E.T]\right\}$$

$$\frac{d}{dt}[E] = \frac{d}{dt}[DNA] + \frac{d}{dt}[T_i] - k_1^e[E]\{[S_1P_1] + [S_2P_2]\} + k_{-1}^e\{[E.S_1P_1] + [E.S_2P_2]\}$$

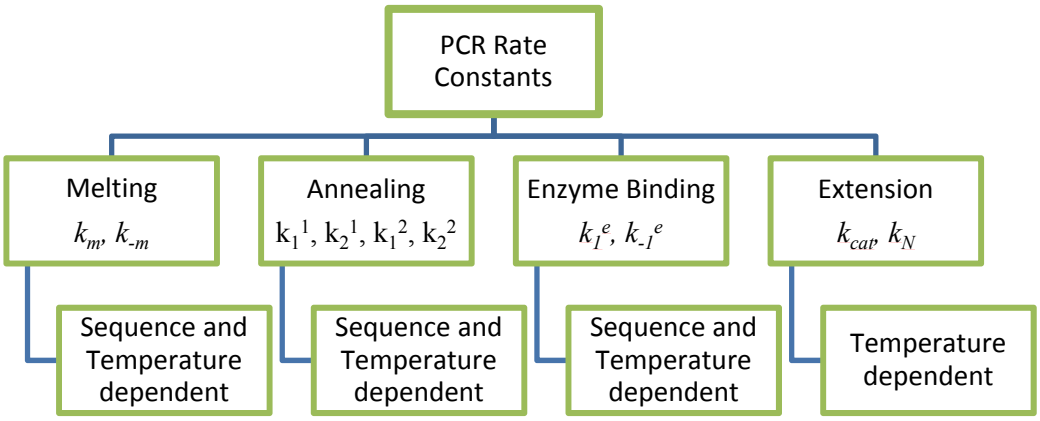


Figure A1: Sequence and Temperature dependent PCR Model (Rate Constants)

References:

1. Cobbs, G. 2012. Stepwise kinetic equilibrium models of quantitative polymerase chain reaction. BMC bioinformatics 13:203.
2. Chakrabarti, R., and C. Schutt. 2001. The enhancement of PCR amplification by low molecular-weight sulfones. Gene. 274:2377-2381.
3. Chakrabarti, R., and C. Schutt. 2001. The enhancement of PCR amplification by low molecular weight amides. Nucleic Acids Research. 62:383-401.
4. Skladny, H., D. Buchheidt, C. Baust, F. Krieg-Schneider, W. Seifarth, C. Leib-Mösch, and R. Hehlmann. 1999. Specific detection of *Aspergillus* species in blood and bronchoalveolar lavage samples of immunocompromised patients by two-step PCR. Journal of Clinical Microbiology, *37*(12), 3865-3871.
5. Stengel, R. 1994. Optimal Control and Estimation. New York: Dover.

6. Rychlik, W., W. Spencer, and R. Rhoads. 1990. Optimization of the annealing temperature for DNA amplification in vitro. Nucleic Acids Research. 18:6409-6412.

7. Stolovitzky, G., and G. Cecchi. 1996. Efficiency of DNA replication in the polymerase chain reaction. Proceedings of National Academy of Sciences, USA. 93:12947-12952.

8. Velikanov, M., and R. Kapral. 1999. Polymerase chain reaction: A Markov process approach. Journal of Theoretical Biology. 201:239-249.

9. Yang, I., Y. Kim, J. Byun, and S. Park. 2005. Use of multiplex polymerase chain reactions to indicate the accuracy of the annealing temperature of thermal cycling. Analytical Biochemistry. 338:192-200.

10. Mehra, S., and W. Hu. 2005. A kinetic model of quantitative real time polymerase chain reaction. Biotechnology and Bioengineering. 91:848-860.

11. Hsu, J., S. Das, and S. Mohapatra. Polymerase chain reaction engineering. 1997. Biotechnology and Bioengineering. 55:359-366.

12. Gevertz, J., S. Dunn, and C. Roth. 2006. A mathematical model of real-time pcr kinetics. Biotechnology and Bioengineering. 92:346-366.

13. Lee, J., H. Lim, S. Yoo, B. Zhang, and T. Park. 2006. Simulation and real-time monitoring of polymerase chain reaction for its higher efficiency. Biochemical Engineering Journal. 29:109-118.

14. Griep, M., S. Whitney, M. Nelson, and H. Viljoen. 2006. DNA polymerase chain reaction: A model of error frequencies and extension rates. AIChE Journal. 52:384-392.

15. Data, K., and V. LiCata. 2003. Thermodynamics of the binding of Thermus aquaticus dna polymerase to primed-template dna. Nucleic Acids Research. 31:5590-5597.

16. Hung, M., N. Arnheim, and M. Goodman. 1992. Extension of base mispairs by Taq DNA polymerase: implications for single nucleotide discrimination in pcr. Nucleic Acids Research. 20: 4567-4573.

17. Innis, M. A., K. Myambo, D. Gelfand, and M. Brow. 1988. DNA sequencing with thermus aquaticus dna polymerase anddirect sequencing of polymerase chain reactionamplified dna. Proceedings of National Academy of Sciences USA. 85:9436-9440.

18. Marimuthu, M., and R. Chakrabarti. 2014. Sequence-dependent oligonucleotide hybridization kinetics. Journal of Chemical Physics. J. Chem. Phys. 140: 175104.

19. Craig, M., D. Crothers, and P. Doty. 1971. Relaxation kinetics of dimer formation by self-complementary oligonucleotides. Journal of Molecular Biology. 62:383-401.

20. Porschke, D., and M. Eigen. 1971. Co-operative non-enzymic base recognition.iii. kinetics of the helix-coil transition of the oligoribouridylic.oligoriboadenylic acid system and of ligoriboadenylic acid alone at acidic ph. Journal of Molecular Biology. 62:361-381.

21. Koehle, R., and N. Peyret. 2005. Thermodynamic properties of DNA sequences: characteristic values for the human genome. Bioinformatics. 21:3333-3339.

22. Garel, T., and H. Orland. 2004. Generalized Poland-Scheraga model for dna hybridization. Biopolymers. 75:453-467.

23. Jost, D., and R. Everaers. 2009. A unified Poland-Scheraga model of oligo and polynucleotide DNA melting: Salt effects and predictive power. Biophysical journal. 96:1056-106.

24. Richard, C., and A. J. Guttmann. 2004. Poland-Scheraga models and the DNA denaturation transition. Journal of statistical physics. 115:925-947.

25. Jin, L., L. Wang, H. Mamon, M. H. Kulke, R. Berbeco, and G. M. Makrigiorgos. Replacing PCR with COLD-PCR enriches variant DNA sequences and redefines the sensitivity of genetic testing. *Nature medicine* 14, no. 5 (2008): 579-584.

26. Blake, R., and S. G. Delcourt. 1998. Thermal stability of DNA. Nucleic acids research. 26:3323-3332.

27. Blake, R., J. Bizzaro, B. J.D, G. Day, S. Delcourt, J. Knowles, K. Marx, and J. Santalucia. 1999. Statistical thermodynamical simulation of polymeric DNA melting with meltsim. Bioinformatics. 15: 370-375.

28. Dwight, Z., R. Palais, and C. T. Wittwer. 2011. Umelt: prediction of high-resolution melting curves and dynamic melting profiles of PCR products in a rich web application. Bioinformatics. 27: 1019-1020.

29. Azbel, M. Y., DNA sequencing and melting curve. 1979. Proceedings of the National Academy of Sciences. 76:101-105.

30. Kuchta, R., V. Mizrahi, P. Benkovic, K. Johnson, and S. Benkovic. 1987. Kinetic mechanism of DNA polymerase i (klenow). Biochemistry. 26:8410-8417.

31. Patel, S. S., I. Wong, and K. A. Johnson. 1991. Pre-steady-state kinetic analysis of processive DNA replication including complete characterization of an exonuclease-deficient mutant. Biochemistry. 30:511-525.

32. Boosalis, M., J. Petruska, and M. Goodman. 1987. DNA polymerase insertion fidelity. The Journal of Biological Chemistry. 262:14689-14696.

33. Mendelman, L. V., J. Petruska, and M. F. Goodman. 1990. Base mispair extension kinetics. Comparison of DNA polymerase alpha and reverse transcriptase. Journal of Biological Chemistry. 265:2338-2346.

34. Huang, M. M., N. Arnheim, and M. F. Goodman. 1992. Extension of base mispairs by taq dna polymerase: implications for single nucleotide discrimination in PCR. Nucleic acids research. 20: 4567-4573.

35. Brown, J. A., and Z. Suo. 2009. Elucidating the kinetic mechanism of DNA polymerization catalyzed by *Sulfolobus solfataricus* P2 DNA Polymerase B1. Biochemistry. 48: 7502-7511.

36. Capson, T. L., J. A. Peliska, B. F. Kaboord, M. W. Frey, C. Lively, M. Dahlberg, and S. J. Benkovic. 1992. Kinetic Characterization of the polymerase and exonuclease activities of the gene 43 protein of bacteriophage T4. Biochemistry. 31:10984-10994.

37. Fiala, K. A., and Z. Suo. 2004. Mechanism of DNA polymerization catalyzed by Sulfolobus solfataricus P2 DNA polymerase IV. Biochemistry, *43*(7), 2116-2125.

38. Wang, Y., D.E. Prosen, L. Mei, J.C. Sullivan, M. Finney, and P.B. Vander Horn. 2004. A novel strategy to engineer DNA polymerases for enhanced processivity and improved performance in vitro. Nucleic acids research, *32*(3), 1197-1207.

39. Davidson, J. F., R. Fox, D.D. Harris, S. Lyons-Abbott, and L. A. Loeb. 2003. Insertion of the T3 DNA polymerase thioredoxin binding domain enhances the processivity and fidelity of Taq DNA polymerase. Nucleic acids research, *31*(16), 4702-4709.

40. Invitrogen, Pyrophosphate assay kit. 2006. http://probes.invitrogen.com/media/pis/mp22062.pdf.