12-3-2007

Mathematically Modeling PCR: An Asymptotic Approximation with Potential for Optimization

Martha J. Garlick
Utah State University

Mathematically Modeling PCR:
An Asymptotic Approximation with Potential for Optimization

By Martha J. Garlick

A report submitted in partial fulfillment of the requirements for the degree of Master of Science in Mathematics

Utah State University
Logan, Utah
2007

Mathematically Modeling PCR:

An Asymptotic Approximation with Potential for Optimization

Martha J. Garlick

December 3, 2007

Abstract

A mathematical model for PCR (Polymerase Chain Reaction) is developed using the law of mass action. Differential equations are written from the chemical equations, preserving the detail of the complementary DNA single strand being extended one base pair at a time. The equations for the annealing stage are solved analytically. The method of multiple scales is used to approximate solutions for the extension stage. A map is then developed from the solutions to simulate PCR. The advantage of this model is the ability to use the map to optimize the process. Our results suggest that dynamically optimizing the extension and annealing stages may significantly reduce the total time for a PCR run.

1 Introduction

The Polymerase Chain Reaction (PCR) is a technique for enzymatic amplification of specific segments of DNA. Since its inception (Saiki et al., 1985), it has revolutionized research involving genomic material. Pathogen detection, disease diagnosis, human genetics and developmental biology are just a few of the research areas impacted by PCR (Stone et al., 2006).

PCR is performed by repeating three temperature-induced stages: dissociation, annealing, and

extension. In dissociation a sample containing the target DNA is first heated to approximately 95^oC to separate the DNA into single strands. The mixture is then cooled to allow primers to anneal to the template DNA. Primers are short single strands of DNA specifically designed to target and bracket the sequence of DNA in the sample to be duplicated (the amplicon). The temperature of this stage is primer-specific, ranging from 37^oC to 72^oC. The solution is then heated to 74^oC for extension. During this phase the thermostable enzyme Taq Polymerase synthesizes a new DNA strand, completing the complimentary sequence started by the primer. These three stages are repeated 30 to 40 times yielding millions of copies of the target DNA.

Real-time PCR uses fluorescent probes to monitor the amplification of DNA throughout the reaction. The speed at which the fluorescent signal reaches a threshold level correlates with the amount of target DNA in the initial sample. Real-time PCR is used to precisely distinguish and measure specific DNA sequences even if there is only a very small quantity present in the original sample (Valasek and Repa, 2005). This technology has many applications, including those that benefit from rapidity. Identification of microbes or parasites in commercial food and municipal water supplies, pathogen detection, and forensic applications are just a few. Portable, rapid, real-time PCR machines can determine the presence of a pathogen, such as anthrax, in as little as 30 minutes. However, 30 minutes can be a long time on a battlefield or in the event of a biological terror attack. Using mathematics to optimize the process to potentially reduce this time is a valuable exercise.

PCR has been mathematically modeled in several different ways. Early models assumed growth per cycle proportional to the amplicon density (Stone et al., 2006). However, assuming exponential growth of template copies greatly oversimplifies the process and models the growth only for the first few cycles. In reality, limiting factors cause the process to slow and eventually stop. These factors may include exhaustion of primer molecules and raw base pairs or a decrease in the effectiveness

of Taq (Liu and Saint, 2002).

In consideration of this limiting behavior, several mathematical models in the literature predict the efficiency of the PCR process. Liu and Saint (2002) used a sigmoidal mathematical model to fit real time PCR data, and demonstrated that amplification efficiency can change from cycle to cycle. A linear regression approach to calculating PCR efficiency was given by Ramakers et al. (2002), and Rutledge and Côté (2003) proposed a simplified method for absolute quantification. Gevertz et al. (2005) considered the efficiency of the reaction as a function of the cycle number. They considered an equilibrium model as well as a kinetic model by deriving differential equations directly from the chemical equations for the annealing and extension phases. Aach and Church (2003) also derived mathematical equations from the chemical reactions, but for diffusion-constrained PCR reactions.

Stochastic and probabilistic models are also used to model PCR. Velikanow and Kapral (1999) treated the extension step as a microscopic Markov process in which the nucleotides bind onto the primed single strand of DNA one at a time. Sun et al. (1996) used the theory of branching processes to develop a model for distributions of mutations and estimation of mutation rates during PCR. Weiss and Von Hessler (1995) treated the accumulation of new molecules during PCR as a random bifurcation tree to estimate overall error rates for the reaction. More recently, Jagers, et al. (2003) used Galton-Watson branching processes to arrive at a linear growth phase following the initial exponential phase. Lalam (2006) based another model on a Galton-Watson branching process description of PCR to estimate the reaction efficiency. A drawback of many of these models, particularly from the standpoint of optimization, is their complexity, which requires numerical integration and obscures dynamical understanding of the process.

The model we present uses the chemical reactions of PCR to derive a system of differential equations, mimicking the physical behavior of the single-stranded DNA (ssDNA) copy being extended one nucleotide base pair at a time. Addition of individual nucleotides occurs very rapidly,

and the amount of Taq is small relative to both primer and base pair concentration. This provides leverage to apply an asymptotic solution strategy. There are two main objectives for the solution strategy for this mathematical model. First, the approximate solutions found by a multiple time scale approach can be put into a map to simulate the PCR process. Second, the solutions and the map can be used to optimize the time spent in each stage in order to obtain the most amplification in the shortest possible time. Our results indicate that dynamic optimizing of the extension and annealing phases may significantly decrease the time required for the entire PCR process.

2 The Chemical Reactions of PCR

2.1 Dissociation

Dissociation occurs when the sample containing the double-stranded DNA (dsDNA) is heated to separate the dsDNA into single strands. The chemical equation can be written as

$$D \rightarrow 2S,$$

where D is dsDNA and S is single-stranded DNA (ssDNA). Experimentation has shown that dsDNA held at temperatures above 94^{o}C for more that five seconds is completely denatured (Gevertz et al., 2005). This justifies the assumption that dissociation is complete. Using lower case letters to indicate concentrations ($s = [S], d = [D]$), we represent this stage mathematically by $s = 2d$.

2.2 Annealing

After the dsDNA is denatured, two complementary ssDNA templates are formed. A primer is designed to anneal at the end of the target DNA template for each of the two complementary strands. We simplify the reaction by including only one chemical equation for this, assuming that the priming for each of the complementary strands occurs at the same rate. The chemical equation

for annealing describes the primer, P, attaching to the ssDNA, S, to form a molecule of primed ssDNA, S'. It can be written as

$$S + P \overset{k_1, k_{-1}}{\longleftrightarrow} S'.$$

The constant k_1 is the rate the reaction moves forward, creating primed ssDNA. A constant k_{-1} is included to model the reverse reaction of primers falling off of previously primed ssDNA. Reaction temperatures are chosen so that $k_1 >> k_{-1}$, allowing the reaction to proceed rapidly.

Other reactions can occur during annealing. The two complementary template strands can reanneal and primers can anneal to each other instead of to the ssDNA. Since the length of the strand is generally greater than 100 base pairs, reannealing of the complementary templates is unlikely and we assume it does not happen. We also neglect the scenario of primer to primer annealing, since tremendous effort is exerted in primer design to prevent this (Eyre, 2005).

2.3 Extension

In the extension stage, Taq polymerase, Q, binds with primed ssDNA, S', to form a complex, C, at the rate λ_1,

$$S' + Q \overset{\lambda_1}{\rightarrow} C.$$

Taq facilitates the addition of base pairs in order from the primer to the end of the strand. We write a separate equation for each base pair added. C_j denotes the complex with j base pairs ($j = 1, ..., n$). The number, n, denotes the number of nucleotide base pairs needed to complete the complementary strand and R represents the resources containing all 4 types of individual base pairs for extension. We assume that all base pairs add on to the template at the same rate, λ_2, and that all are present and needed in equal proportions.

The equations for this process are as follows:

$$C + R \overset{\lambda_2}{\rightarrow} C_1,$$

$$C_1 + R \xrightarrow{\lambda_2} C_2,$$

$$\vdots$$

$$C_{n-1} + R \xrightarrow{\lambda_2} C_n.$$

The Taq separates from the dsDNA as the template copy is completed at the rate λ_3,

$$C_n \xrightarrow{\lambda_3} D + Q.$$

At the reaction temperatures used for PCR, Taq is quite efficient at synthesizing the complementary strands. Therefore, we consider any back reactions in this stage to be negligible. Descriptions of the reactants and constants are shown in Table 1.

Symbols	Description
D, d	double stranded DNA (dsDNA), $d = [D]$ (concentration)
S, s	single stranded DNA (ssDNA), $s = [S]$
P, p	primer, $p = [P]$
S', s'	primed ssDNA, $s' = [S']$
Q, q	Taq polymerase, $q = [Q]$
C, c	complex of primed ssDNA with Taq, $c = [C]$
C_j, c_j	complex with j base pairs added, $j = 1, 2, ..., n, c_j = [C_j]$
R, r	resources containing nucleotides for extension, $r = [R]$
k_1, k_{-1}	forward and backward reaction rates for annealing
$\lambda_1, \lambda_2, \lambda_3$	forward reaction rates for extension stage
$\hat{s}$	initial amount of ssDNA for a single cycle
$\hat{p}$	initial amount of primer for a single cycle
$\hat{s}'$	initial amount of primed ssDNA for a single cycle
$\hat{q}$	initial amount of Taq for a single cycle
$\hat{r}$	initial amount of resources containing the individual base pairs
τ	rescaled time for extension stage ($\lambda_2 \hat{r} t$)
$\bar{s}$, $\bar{q}$, etc.	rescaled reactants for extension stage
ϵ	$\frac{\lambda_1 \hat{q}}{\lambda_2 \hat{r}}$
ν	$\frac{\lambda_1 \hat{p}}{\lambda_2 \hat{r}}$
μ	$\frac{\lambda_3}{\lambda_2 \hat{r}}$

Table 1: List of variables and constants used in PCR model.

3 Model Development

3.1 Annealing

The law of mass action is invoked to write a system of differential equations for the annealing stage of a single cycle of PCR. This approach is particularly justified in the case of quantitative PCR where reactions occur in small, well-mixed containers.

$$\frac{ds}{dt} = -k_1 sp + k_{-1} s', \tag{1}$$

$$\frac{dp}{dt} = -k_1 sp + k_{-1} s', \tag{2}$$

and

$$\frac{ds'}{dt} = k_1 sp - k_{-1} s'. \tag{3}$$

Equation (1) represents the change in the concentration of ssDNA as a function of the concentrations of ssDNA s, primer, p, and primed ssDNA, s', scaled by the forward and backward reaction rates, k_1 and k_{-1}. Likewise, equations (2) and (3) describe the change in concentration of primer and primed ssDNA. The initial conditions are $s'(0) = 0$ (since no primed ssDNA survives denaturing), $s(0) = \hat{s}$ (the amount of ssDNA after dissociation), and $p(0) = \hat{p}$ (the amount of primer at the beginning of this stage). It can be observed from equations (1)-(3) that $\frac{ds}{dt} + \frac{ds'}{dt} = 0$ and $\frac{dp}{dt} + \frac{ds'}{dt} = 0$. This gives rise to two conserved quantities:

$$s + s' = K_1 = s(0) + s'(0) = s(0) = \hat{s}$$

and

$$p + s' = K_2 = p(0) + s'(0) = p(0) = \hat{p}.$$

Using these quantities the system can be reduced to a single equation,

$$\frac{ds'}{dt} = k_1(\hat{p} - s')(\hat{s} - s') - k_{-1} s', \tag{4}$$

and solved analytically, using separation of variables. The solution is

$$s'(t) = \frac{\alpha_+\alpha_-(1 - e^{(\alpha_- - \alpha_+)t})}{\alpha_+ - \alpha_- e^{(\alpha_- - \alpha_+)t}}, \tag{5}$$

where

$$\alpha_\pm = \frac{k_{-1} + k_1\hat{p} + k_1\hat{s} \pm \sqrt{(k_{-1} + k_1\hat{p} + k_1\hat{s})^2 - 4k_1^2\hat{p}\hat{s}}}{2k_1}. \tag{6}$$

The sigmoidal solution (5) increases to a limiting quantity determined by the initial amounts of ssDNA and primer.

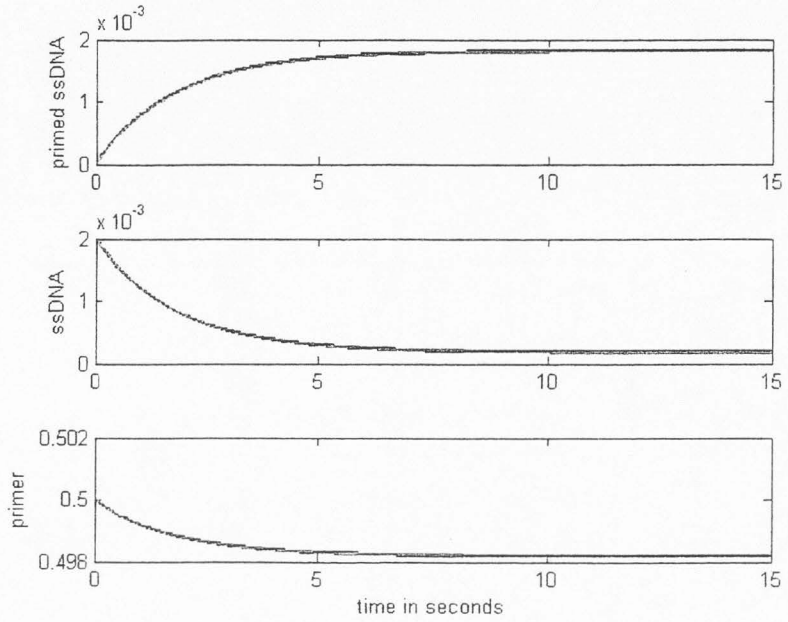

Figure 1: The graphs of the exact solutions for the annealing stage of a single cycle of PCR with $\hat{p} = .5$ and $\hat{s} = .002$. The rate constants used are $k_1 = 0.205$ and $k_{-1} = 0.01025$. The solution curve for primed ssDNA levels off when all of the ssDNA strands have been primed.

The solutions for primer and single-stranded DNA come from the conserved quantities and are:

$$s(t) = \hat{s} - s'(t) \tag{7}$$

and

$$p(t) = \hat{p} - s'(t). \tag{8}$$

A graph of this solution is shown in Figure (1).

3.2 Extension

The law of mass action is again applied to write differential equations (9)-(16) that describe the change in concentration of each of the reactants as Taq extends the template copy. The constants, λ_1,λ_2, and λ_3 are forward reaction rates for this stage. The change in the concentration of primed ssDNA, is proportional to the product of the concentrations of primed ssDNA, s', and unattached Taq, q, as shown in the equation,

$$\frac{ds'}{dt} = -\lambda_1 s' q. \tag{9}$$

The change in the concentration of unattached Taq is also a function of s' and q as well as the concentration of the complex with all the base pairs added, c_n. This models Taq binding with the primed ssDNA at the beginning of extension and detaching after the template strand is completed giving,

$$\frac{dq}{dt} = -\lambda_1 s' q + \lambda_3 c_n. \tag{10}$$

The change in the concentration of the complex, c, is a function of s', q, c, and the concentration of resources, r,

$$\frac{dc}{dt} = \lambda_1 s' q - \lambda_2 r c. \tag{11}$$

The next $n-1$ equations, represented by (12) and (13), exhibit a distinct pattern as they model addition of base pairs to the template strand. The pattern models the creation of the complex, C_j as the jth base pair is added to the previous complex, C_{j-1}, and its subsequent disappearance when the next base pair is added. The change in the concentration for a particular complex with j base pairs added is a function of the concentration of that complex, c_j, r, and the concentration of the previously formed complex, c_{j-1}.

$$\frac{dc_1}{dt} = \lambda_2 r c - \lambda_2 r c_1 \tag{12}$$

$$\vdots$$

$$\frac{dc_{n-1}}{dt} = \lambda_2 r c_{n-2} - \lambda_2 r c_{n-1}. \tag{13}$$

The equation for the change in the concentration of the complex with n base pairs differs from (12)-(13) and is written,

$$\frac{dc_n}{dt} = \lambda_2 c_{n-1} r - \lambda_3 c_n. \tag{14}$$

It is a function of the concentrations of the previous complex, c_{n-1}, r, and itself, c_n, but it includes λ_3, the rate at which Taq detaches from the completed complex, C_n, forming dsDNA. The change in the concentration of resources is affected by the concentrations of all the complexes up to c_{n-1} as shown in the equation,

$$\frac{dr}{dt} = -\lambda_2 r \sum_{j=1}^{n-1} c_j. \tag{15}$$

The change in the concentration of dsDNA is proportional to the concentration of the complex with all of the base pairs added, c_n:

$$\frac{dd}{dt} = \lambda_3 c_n. \tag{16}$$

The equation for double-stranded DNA, (16), is coupled only to (14) and can be solved by direct integration after (9)-(15) are solved. The initial conditions are: $s'(0) = \hat{s}'$ (the amount of primed ssDNA present at the end of the previous annealing stage), $q(0) = \hat{q}$ (the initial amount of Taq), $c(0) = c_j(0) = d(0) = 0$ (since dissociation peels off any partially competed amplicon), and $r(0) = \hat{r}$ (the amount of resources remaining after the previous extension stage).

An inherent small quantity in this stage of the PCR process is the proportion of the initial amount of Taq to the initial amount of resources, $\frac{\hat{q}}{\hat{r}}$, due to the fact that nucleotide base pairs are relatively easier to obtain and used in much larger quantities than Taq. Therefore, we choose a time scale and concentration scales to form a dimensionless system in a way that lets us take advantage of this small quantity. This amounts to assuming that Taq is the rate-limiting quantity. The time

is non-dimensionalized using $\tau = \lambda_2 \hat{r} t$, and the concentrations are normalized using: $\bar{s}' = \frac{s'}{\hat{p}}$, $\bar{q} = \frac{q}{\hat{q}}$, $\bar{c} = \frac{c}{\hat{q}}$, $\bar{c}_j = \frac{c_j}{\hat{q}}$, and , $\bar{r} = \frac{r}{\hat{r}}$. Using dot notation for $\frac{d}{d\tau}$ ($\frac{d\bar{s'}}{d\tau} = \dot{\bar{s}}'$, etc.), the equations (9)-(15) become:

$$\dot{\bar{s}}' = -\frac{\lambda_1 \hat{q}}{\lambda_2 \hat{r}} \bar{s}' \bar{q}, \tag{17}$$

$$\dot{\bar{q}} = -\frac{\lambda_1 \hat{p}}{\lambda_2 \hat{r}} \bar{s}' \bar{q} + \frac{\lambda_3}{\lambda_2 \hat{r}} \bar{c}_n, \tag{18}$$

$$\dot{\bar{c}} = \frac{\lambda_1 \hat{p}}{\lambda_2 \hat{r}} \bar{s}' \bar{q} - \bar{r}\bar{c}, \tag{19}$$

$$\dot{\bar{c}}_1 = \bar{r}\bar{c} - \bar{r}\bar{c}_1, \tag{20}$$

$$\vdots$$

$$\dot{\bar{c}}_{n-1} = \bar{r}\bar{c}_{n-2} - \bar{r}\bar{c}_{n-1}, \tag{21}$$

$$\dot{\bar{c}}_n = \bar{r}\bar{c}_{n-1} - \frac{\lambda_3}{\lambda_2 \hat{r}} \bar{c}_n, \tag{22}$$

and

$$\dot{\bar{r}} = -\frac{\hat{q}}{\hat{r}} \bar{r} \sum_{j=1}^{n-1} \bar{c}_j, \tag{23}$$

with rescaled initial conditions $\bar{s}'(0) = \frac{\hat{s}'}{\hat{p}} = \gamma$, $\bar{q}(0) = \frac{\hat{q}}{\hat{q}} = 1$, $\bar{r}(0) = \frac{\hat{r}}{\hat{r}} = 1$,and $\bar{c}(0) = \bar{c}_j(0) = 0$.

The quantity $\frac{\lambda_1 \hat{q}}{\lambda_2 \hat{r}}$ in (17), contains the small quantity, $\frac{\hat{q}}{\hat{r}}$. Since the rate constants, λ_1 and λ_2 are of the same order, $\epsilon = \frac{\lambda_1 \hat{q}}{\lambda_2 \hat{r}}$ is small. We define two other dimensionless parameters, $\nu = \frac{\lambda_1 \hat{p}}{\lambda_2 \hat{r}}$, and $\mu = \frac{\lambda_3}{\lambda_2 \hat{r}}$ for simplification. The system (17-(23) becomes:

$$\dot{\bar{s}}' = -\epsilon \bar{s}' \bar{q}, \tag{24}$$

$$\dot{\bar{q}} = -\nu \bar{s}' \bar{q} + \mu \bar{c}_n, \tag{25}$$

$$\dot{\bar{c}} = \nu \bar{s}' \bar{q} - \bar{r}\bar{c}, \tag{26}$$

$$\dot{\bar{c}}_1 = \bar{r}\bar{c} - \bar{r}\bar{c}_1 \tag{27}$$

$$\vdots$$

$$\dot{\bar{c}}_{n-1} = \bar{r}\bar{c}_{n-2} - \bar{r}\bar{c}_{n-1}, \tag{28}$$

$$\dot{\bar{c}}_n = \bar{r}\bar{c}_{n-1} - \mu\bar{c}_n, \tag{29}$$

and

$$\dot{\bar{r}} = -\epsilon\frac{\lambda_2}{\lambda_1}\bar{r}\sum_{j=1}^{n-1}\bar{c}_j. \tag{30}$$

Symbol	Units	Value
$\hat{p}$	$\mu mol\ \mu l^{-1}$	0.5
$\hat{q}$	$\mu mol\ \mu l^{-1}$	0.01
$\hat{r}$	$\mu mol\ \mu l^{-1}$	10
k_1	$\mu l\ \mu mol^{-1}\ sec^{-1}$	0.205
k_{-1}	sec_{-1}	0
λ_1	$\mu l\ \mu mol^{-1}\ sec^{-1}$	9
λ_2	$\mu l\ \mu mol^{-1}\ sec^{-1}$	10
λ_3	sec_{-1}	100

Table 2: Parameters in PCR model and their estimated values in simulations. Concentrations are in micromoles per microliter ($\mu mol\ \mu l^{-1}$) and time is in seconds.

3.3 Multiple Time Scale Analysis

The presence of a small parameter, ϵ, allows the application of the method of multiple time scales. A more detailed description of this method is found in Holmes (1995). Assigning $t_1 = \tau$ to be the fast time scale and $t_2 = \epsilon\tau$ to be the slow time scale, $\frac{d}{d\tau}$ becomes $\frac{\partial}{\partial t_1} + \epsilon\frac{\partial}{\partial t_2} = \partial_{t_1} + \epsilon\partial_{t_2}$. We substitute this new time derivative into equations (24)-(30), along with a power series expansion of the form,

$$y = y_0(t_1, t_2) + \epsilon y_1(t_1, t_2) + ..., \tag{31}$$

for each concentration variable. For example, (24) becomes

$$(\partial_{t_1} + \epsilon\partial_{t_2})(\bar{s}_0' + \epsilon\bar{s}_1' + ...) = -\epsilon(\bar{s}_0' + \epsilon\bar{s}_1' + ...)(\bar{q}_0 + \epsilon\bar{q}_1 + ...). \tag{32}$$

Collecting the order ϵ^0 terms gives the leading order equations:

$$\partial_{t_1}\bar{s}_0' = 0, \tag{33}$$

$$\partial_{t_1}\bar{q}_0 = -\nu\bar{s}_0'\bar{q}_0 + \mu\bar{c}_{n_0}, \tag{34}$$

$$\partial_{t_1}\bar{c}_0 = \nu\bar{s}_0'\bar{q}_0 - \bar{r}_0\bar{c}_0, \tag{35}$$

$$\partial_{t_1}\bar{c}_{1_0} = \bar{r}_0\bar{c}_0 - \bar{r}_0\bar{c}_{1_0}, \tag{36}$$

$$\partial_{t_1}\bar{c}_{2_0} = \bar{r}_0\bar{c}_{1_0} - \bar{r}_0\bar{c}_{2_0}, \tag{37}$$

$$\vdots$$

$$\partial_{t_1}\bar{c}_{n-1_0} = \bar{r}_0\bar{c}_{n-2_0} - \bar{r}_0\bar{c}_{n-1_0}, \tag{38}$$

$$\partial_{t_1}\bar{c}_{n_0} = \bar{r}_0\bar{c}_{n-1_0} - \mu\bar{c}_{n_0}, \tag{39}$$

and

$$\partial_{t_1}\bar{r}_0 = 0. \tag{40}$$

The initial conditions become: $\bar{s}_0'(t_1 = t_2 = 0) = \gamma$, $\bar{q}_0(t_1 = t_2 = 0) = \bar{r}_0(t_1 = t_2 = 0) = 1$, and $\bar{c}_0(t_1 = t_2 = 0) = \bar{c}_{j_0}(t_1 = t_2 = 0) = 0$.

Equations (33) and (40) imply that $\bar{r}_0$ and $\bar{s}_0'$ are constant on the fast time scale. This means that the system of equations,(34)-(39), is linear and can be written in vector form as

$$\boldsymbol{x}'(t_1) = A\boldsymbol{x}(t_1),$$

where A is the coefficient matrix. The sum of the rows of A equals zero, and it is easy to show that A has one zero eigenvalue and $n+1$ eigenvalues less than zero. Therefore, the solution for this linear system takes the form,

$$\boldsymbol{x}(t_1) = \boldsymbol{v}_0 + \sum_{j=1}^{n} e^{\lambda_j t_1}\boldsymbol{v}_j,$$

where $\boldsymbol{v}_0$ is an eigenvector associated with the zero eigenvalue, and the λ_j's are the negative eigenvalues, with their associated eigenvectors, $\boldsymbol{v}_j$. Consequently, $\boldsymbol{x}(t_1) \to \boldsymbol{v}_0$ exponentially fast on the fast time scale. Ignoring these transients, the leading order solution becomes $\boldsymbol{x}(t_1) = \boldsymbol{v}_0$.

With the above in mind, we determine $\boldsymbol{v}_0$ using the conserved quantity obtained by adding together the right sides of (34)-(39),

$$\bar{q}_0 + \bar{c}_0 + \sum_{j=1}^{n-1} \bar{c}_{j_0} = K = \bar{q}_0(0) + \bar{c}_0(0) + \sum_{j=1}^{n-1} \bar{c}_{j_0}(0) = 1. \tag{41}$$

$K = 1$ is determined using the initial conditions. Solving $\bar{q}_0 + \bar{c}_0 + \sum_{j=1}^{n-1} \bar{c}_{j_0} = 1$ for $\bar{q}_0$ yields,

$$\bar{q}_0 = 1 - \bar{c}_0 - \sum_{j=1}^{n-1} \bar{c}_{j_0}. \tag{42}$$

Setting the right hand sides of (35)-(39) to 0 and solving yields:

$$\bar{c}_{n_0} = \frac{\nu \bar{s}_0'}{\mu} \bar{q}_0, \tag{43}$$

$$\bar{c}_0 = \nu \bar{s}_0' \bar{q}_0, \tag{44}$$

and

$$\bar{c}_0 = \bar{c}_{1_0} = \bar{c}_{2_0} = \cdots = \bar{c}_{n-1_0}. \tag{45}$$

Using (43)-(45), the equation for $\bar{q}_0$, (42), becomes

$$\bar{q}_0 = \frac{\mu}{\mu + n\mu\nu\bar{s}_0' + \nu\bar{s}_0'}, \tag{46}$$

where n, the number of base pairs added, acts as a shape parameter.

The order ϵ^1 equation for $\bar{s}'$ is

$$\partial_{t_1} \bar{s}_1' = -\bar{s}_0' \bar{q}_0 - \partial_{t_2} \bar{s}_0'. \tag{47}$$

Solving this equation yields:

$$\bar{s}_1' = -(\bar{s}_0' \bar{q}_0 + \partial_{t_2} \bar{s}_0') t_1, \tag{48}$$

since $\bar{s}_0'$ and $\bar{q}_0$ are functions only at t_2. To eliminate the secular term in (48), we require

$$0 = \partial_{t_2} \bar{s}_0' + \bar{s}_0' \bar{q}_0 = \partial_{t_2} \bar{s}_0' + \frac{\mu \bar{s}_0'}{\mu + n\mu\nu\bar{s}_0' + \nu\bar{s}_0'} = 0, \tag{49}$$

using the expression for $\bar{q}_0$ in (46). We solve (49) using separation of variables. Separating and integrating gives

$$\mu \ln \bar{s}_0' + (n\mu + 1)\nu \bar{s}_0' = -\mu t_2 + K, \tag{50}$$

where K is an integration constant. The initial condition $\bar{s}_0'(t_2 = 0) = \gamma$ gives

$$K = \mu \ln \gamma + (n\mu + 1)\nu\gamma.$$

The implicit solution for s', (50), becomes

$$\mu \ln \bar{s}_0' + (n\mu + 1)\nu \bar{s}_0' = -\mu t_2 + \mu \ln \gamma + (n\mu + 1)\nu\gamma,$$

which can be solved for t_2,

$$t_2 = \frac{\mu \ln \bar{s}_0' - (n\mu + 1)\nu \bar{s}_0' + \mu \ln \gamma + (n\mu + 1)\nu\gamma}{\mu}. \tag{51}$$

Returning to the original scale, equations (51) and (43)-(46) become:

$$t = \frac{\ln \frac{\hat{s'}}{s'}}{\epsilon \lambda_2 \hat{r}} + \frac{\nu(n\mu + 1)(\hat{s'} - s')}{\mu \epsilon \lambda_2 \hat{r} \hat{p}}, \tag{52}$$

$$q = \frac{\mu \hat{q} \hat{p}}{\mu \hat{p} + n\mu\nu s' + \nu s'}, \tag{53}$$

$$c_n = \frac{\nu \hat{q} s'}{\mu \hat{p} + n\mu\nu s' + \nu s'}, \tag{54}$$

$$c = c_1 = c_2 = \cdots = c_{n-1} = \mu c_n. \tag{55}$$

The solution to our original equation for double-stranded DNA is

$$d = \lambda_3 \int_0^{t_{end}} c_n(t) dt. \tag{56}$$

The units used in the original scale are shown in Table 2.

Figure 2 shows a comparison between the asymptotic and numerical solutions for the concentration of dsDNA in the extension stage. A Runge Kutta method was used for the numerical solution. Parameter values were estimated. The asymptotic and numerical solutions converge as ϵ tends towards zero.

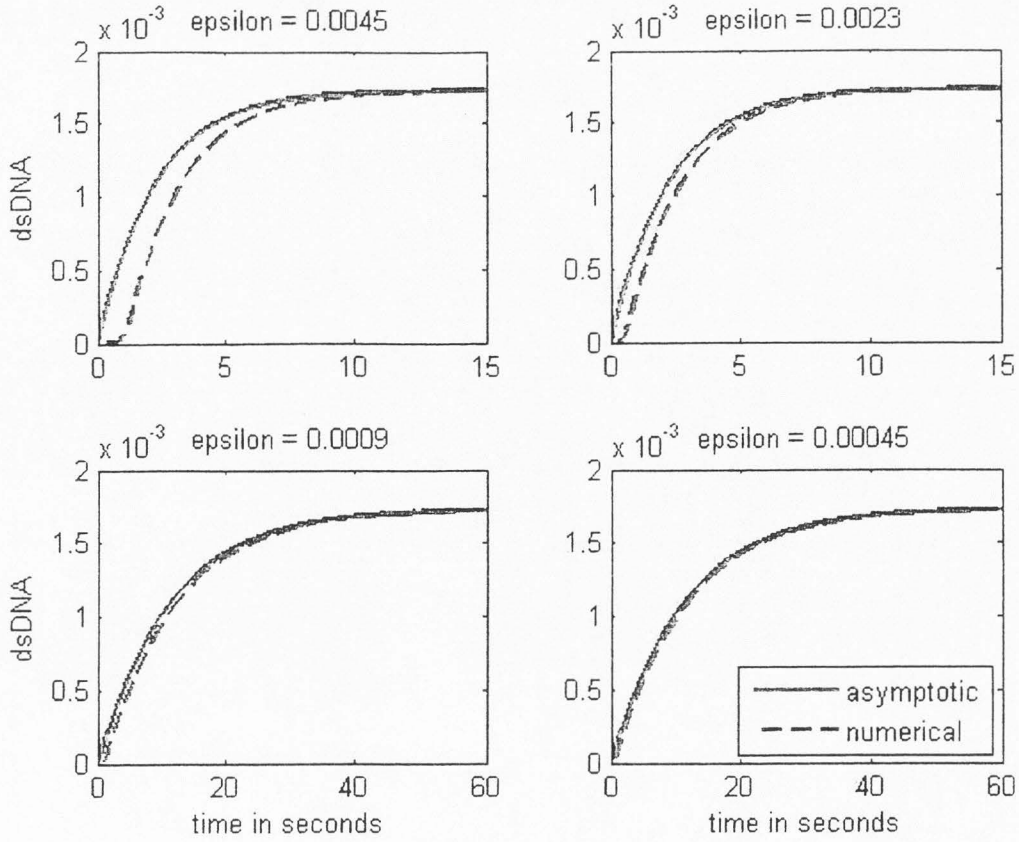

Figure 2: Comparison of asymptotic (solid lines) and numerical (dashed lines) solutions for dsDNA in the extension stage for various values of ϵ with a target strand length of n=100. A Runge Kutta method was used to generate the numerical solutions. As ϵ tends towards zero, the asymptotic and numerical solutions converge.

4 Multi-cycle Map

A map from one cycle of PCR to the next can be constructed from the solutions for the annealing and extension stages and their initial conditions. This is done by cascading the results of a previous cycle into the initial conditions for the next cycle. Let the concentration of primed ssDNA be represented by s'^A for the annealing stage and by s'^E for the extension phase. Also, let the final times for the annealing and extension stages be fixed at t_A and t_E respectively.

The first cycle begins with initial amounts of resources, primer and Taq, and a sample containing the dsDNA to be duplicated. The amount of resources, $\hat{r}$, is considered as essentially constant in

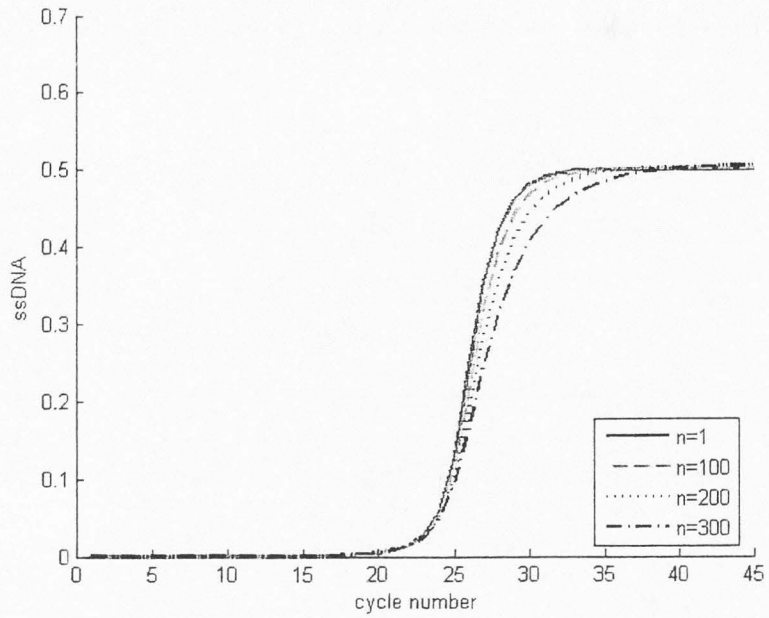

Figure 3: The PCR map for various lengths of target DNA. The longer the target strand, the longer it takes the solution to reach a limiting value. Thus, n acts as a shape parameter. Due to a relationship between ϵ and n in the solutions, the error increases as n increases, causing the solutions to rise above the expected limiting value (the initial amount of primer).

the map; the amount of resource used is too small to have an appreciable effect. Assuming that dissociation is complete, the initial amount of ssDNA for the annealing stage of the first cycle is $s_1(0) = 2d_1$. The other initial conditions for the annealing stage of the first cycle are $p_1(0) = \hat{p}$ (the amount of primer at the beginning of the PCR run) and $s_1'^A(0) = 0$. For all other cycles dissociation not only denatures the dsDNA created in the extension phase, it also denatures any primed ssDNA and any complexes that remain after extension. Thus, for the annealing stage of the ith cycle, the initial amount of ssDNA, $s_i(0)$, includes not only twice the amount of dsDNA from the previous cycle, but also the amount of ssDNA, the amounts of primed ssDNA, and the amount of all of the complexes from the previous cycle. This can be written as

$$s_i(0) = 2d_{i-1}(t_E) + s_{i-1}(t_A) + s_{i-1}'^E(t_E) + c_{i-1}(t_E) + c_{1i-1}(t_E) + \ldots + c_{n-1i-1}(t_E) + c_{ni-1}(t_E).$$

Likewise, the initial amount of primer for the ith cycle is the sum of the amount of primer from

the previous cycle plus the primers gained from denaturing the primed ssDNA and the complex of Taq and primed ssDNA from the previous cycle, written as,

$$p_i(0) = p_{i-1}(t_A) + s'^{E}_{i-1}(t_E) + c_{i-1}(t_E).$$

The initial condition for primed ssDNA is $s'^{A}_i(0) = 0$, because dissociation is assumed to be complete.

Then using solutions (5)-(8), the map for the annealing stage of the ith cycle is:

$$s'^{A}_i(t_A) = \frac{\alpha_+\alpha_-(1 - e^{(\alpha_- - \alpha_+)t_A})}{\alpha_+ - \alpha_- e^{(\alpha_- - \alpha_+)t_A}},$$

where

$$\alpha_{\pm i} = \frac{k_{-1} + k_1 p_i(0) + k_1 s_i(0) \pm \sqrt{(k_{-1} + k_1 p_i(0) + k_1 s_i(0))^2 - 4k_1^2 p_i(0) s_i(0)}}{2k_1},$$

$$s_i(t_A) = s_i(0) - s'^{A}_i(t_A),$$

and

$$p_i(t_A) = p_i(0) - s'^{A}_i(t_A).$$

The amount of primed ssDNA at the beginning of extension stage for the ith cycle is the amount at the end of the annealing stage of that same cycle, $s'^{E}_i(0) = s'^{A}_i(t_A)$. The initial amount of Taq is the beginning amount for the first cycle, $q_1 = \hat{q}$ and the amount from the previous cycle for the rest of the cycles, $q_i(0) = q_{i-1}(t_E)$. The initial conditions for dsDNA and all the complexes for the extension stage of the ith cycle are: $c_i(0) = c_{1i}(0) = c_{2i}(0) = \cdots = c_{ni}(0) = 0$ and $d_i(0) = 0$, because dissociation is assumed to be complete.

A map for the extension stage can be derived from the solutions (51)-(55). The value for $s'^{E}_i(t_E)$ is extracted numerically from the implicit relationship

$$t_E = \frac{\ln \frac{s'^{E}_i(0)}{s'^{E}_i(t_E)}}{\epsilon\lambda_2\hat{r}} + \frac{\nu(n\mu + 1)(s'^{E}_i(0) - s'^{E}_i(t_E))}{\mu\epsilon\lambda_2\hat{r}p_i(0)}.$$

The concentrations for the rest of the reactants are given by:

$$q_i = \frac{\mu q_i(0) p_i(0)}{\mu p_i(0) + n\mu\nu s_i'^E(t_E) + \nu s_i'^E(t_E)},$$

$$c_{ni} = \frac{\nu q_i(0) s_i'^E(t_E)}{\mu p_i(0) + n\mu\nu s_i'^E(t_E) + \nu s_i'^E(t_E)},$$

$$c_i(t_E) = c_{1i}(t_E) = c_{2i}(t_E) = \cdots = c_{n-1i}(t_E) = \mu c_{ni}(t_E),$$

and

$$d_i(t_E) = \int_0^{t_E} \lambda_3 c_{ni}(s) ds.$$

The equation for $d_i(t_E)$ is solved numerically, using trapezoidal quadrature.

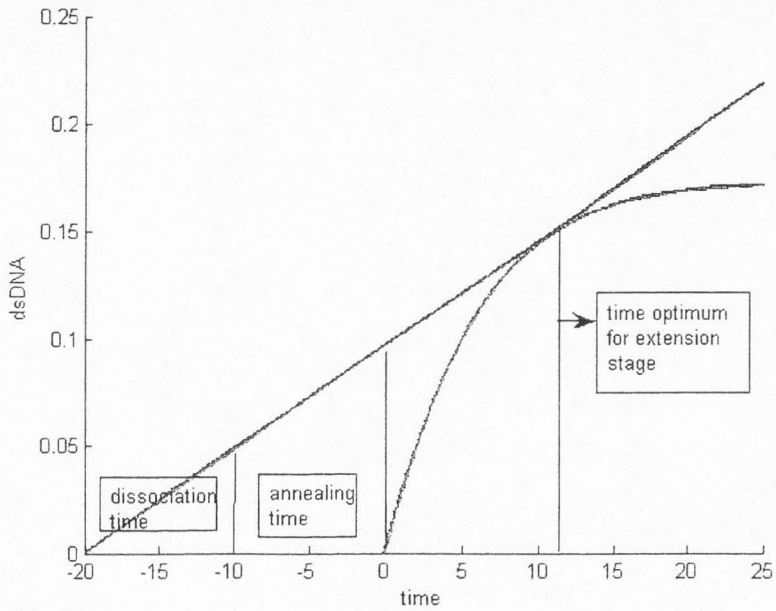

Figure 4: An illustration of the idea of optimization. The blue graph represents the solution for the extension stage for n=200. The negative part of the time axis represents the time spent in the preceding dissociation and annealing stages. The tangent line touches the blue curve at the point of the optimal time to run the extension phase.

The behavior of the solutions can be explored by using this map to simulate the PCR process. However, the parameters must be estimated or fit from the data. This includes t_A, t_E, the rate constants and the initial concentrations of primer, dsDNA, Taq, and resources. One example of

implementing this map with estimated parameters is shown in Figure 3. The lag phase before the solution rises is shorter than it is in most actual PCR runs. One explanation for this is the sensitivity this model has for target DNA strand length. In the early cycles of PCR the amplicon is part of a much longer DNA strand. Until the primers bracket the target sequence, longer strands are being duplicated, which will vastly lower the initial efficiency of the reaction. This map assumes a uniform strand length being amplified from the very first cycle.

5 Optimization

This model and its map may be used to optimize the time spent in each stage of a PCR cycle in order to produce the most DNA copies in the shortest amount of time possible. For an example of how this might be done, we consider optimizing the time spent in the extension stage of a cycle. Let t_D and t_A be the fixed times for the dissociation and annealing stages, respectively. Let t be the variable representing the time spent in the extension stage. Then the total time to complete one cycle of PCR is $t_D + t_A + t$. Let $d(t)$ be the total amount of dsDNA produced in a PCR cycle. Then the t such that

$$\frac{d}{dt}\left(\frac{d(t)}{t_D + t_A + t}\right) = 0,$$

is the optimal time for the extension stage for one cycle. Differentiating,

$$\frac{d}{dt}\left(\frac{d(t)}{t_D + t_A + t}\right) = \frac{d'(t)}{t_D + t_A + t} - \frac{d(t)}{(t_D + t_A + t)^2}. \tag{57}$$

Figure 5 illustrates the idea of optimizing the time spent in the extension stage. Setting the right hand side of (57) equal to zero gives us

$$d'(t) = \frac{d(t)}{t_D + t_A + t}. \tag{58}$$

Then using the right sides of equations (16) and (56) for $d'(t)$ and $d(t)$,(58) becomes

$$(t_D + t_A + t)\lambda_3 c_n = \lambda_3 \int_0^t c_n(s)ds. \tag{59}$$

After simplification we have

$$(t_D + t_A + t)c_n = \int_0^t c_n(s)ds. \tag{60}$$

The t that makes (60) true is found numerically by using the solutions for the extension stage. In order to produce the most dsDNA over the shortest amount of time for an entire PCR run, we include the optimization for the extension stage in the map. The optimal time is calculated and used for each iteration of the extension stage. A set time is used for all the iterations of the annealing stage.

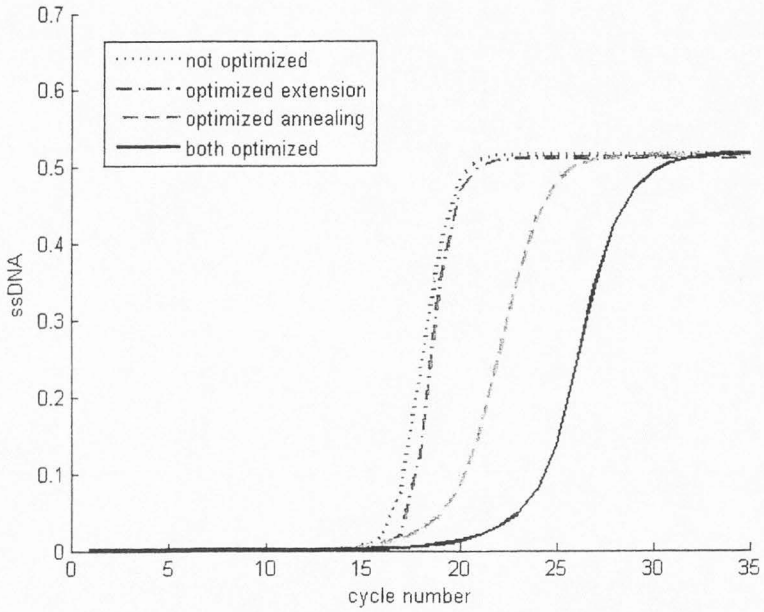

Figure 5: A comparison of optimized runs to a run with a fixed annealing and extension stage times of 10 and 20 seconds, respectively in terms of cycle number. All runs are for target strands of n=100. Even though the optimized runs need more cycles to reach equilibrium, the total time required for the cycles is less.

A similar optimization can be performed for the annealing stage. Figure 5 compares the number of cycles needed for a run with a set time of 10 seconds for annealing and 20 seconds for extension

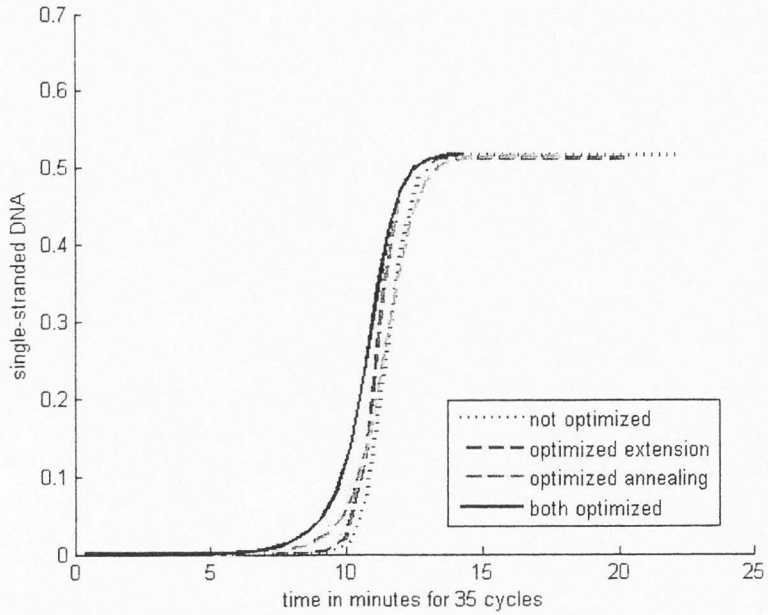

Figure 6: A comparison of optimized runs to a run with a fixed annealing and extension stage times of 10 and 20 seconds, respectively in minutes needed for 35 cycles. Each run has a target strand length of n=100. The total time for 35 cycles is 22 minutes for the non optimized run, 20 minutes for the optimized extension stage run, 18 minutes for the optimized annealing stage run, and 14 minutes when both stages are optimized.

with three optimized runs. First just the extension stage is optimized, second, just the annealing stage, and then both. An 8 second dissociation time is assumed for all runs. The optimized simulations take more cycles to produce the maximum amount of product, but less time as shown in Figure 6. The total time for 35 cycles is 22 minutes for the non optimized run, 20 minutes for the optimized extension stage run, 18 minutes for the optimized annealing stage run, and 14 minutes when both stages are optimized. These results show that optimizing the annealing stage may reduce the time for a PCR run more than just optimizing the extension stage. The time spent in the annealing and extension stages for each cycle for a run with both stages optimized is shown in figure 7.

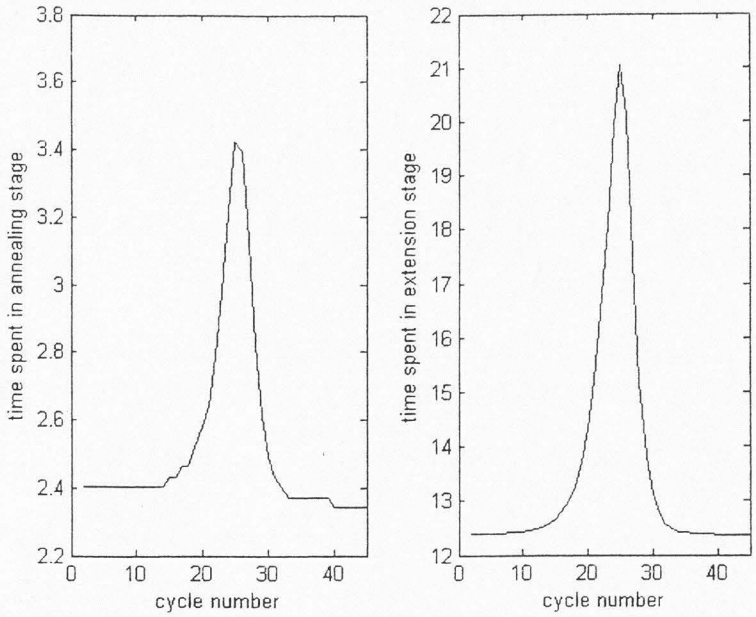

Figure 7: The time spent in the annealing and extension stages for each cycle for a run with both stages optimized. The optimal time changes dynamically with the amount of product being produced. This differs from the set constant times for the annealing and extension stages of the not optimized run.

6 Discussion and Conclusion

In this paper, we have described PCR, discussed existing models, and developed a model from the chemical equations, using the law of mass action. This model of PCR is sensitive to the length of the target DNA strand and models the effect of strand length on the solution shape. We found a simple solution for the dissociation stage, analytical solutions for the annealing stage, and asymptotic approximations for the extension stage, using the method of multiple scales. These solutions were put into a multi-cycle map to simulate PCR. The solutions and the map were then used to optimize the time spent in the extension and annealing stages of each cycle. This research suggests that such an optimization can significantly reduce the overall time for a PCR run.

Further research includes parameterizing the model with PCR data and running optimizations to fit the needs of specific applications. Also, the model could be modified to more realistically

simulate the early cycles when DNA strands longer than the amplicon are being duplicated.

References

Aach, J. and G. M. Church (2004). Mathematical models of diffusion-constrained polymerase chain reactions: basis of high-throughput nucleic acid assays and simple self-organizing systems. J. Theor. Bio. **228**, 31-46.

Eyre, D., private communication, (2005).

Gevertz, J. L., M. D. Stanley, C. M. Roth (2005). Mathematical model of real-time PCR kinetics. Wiley InterScience. (www.interscience.wiley.com) DOI: 10.1002/bit.20617

Holmes, M. H. (1995). Introduction to Perturbation Methods. Springer-Verlag. New York, 105-113.

Jagers, P. and F. Klebaner (2003). Random variation and concentration effects in PCR. J. Theor. Bio. **224**, 299-304.

Lalam, N. (2006). Estimation of the reaction efficiency in polymerase chain reaction. J. Theor. Bio. **242**, 947-953.

Lui, W. and D. A. Saint (2002). Validation of a quantitative method for real tine PCR kinetics. Biochem. Biophys. Res. Com. **294**, 347-353.

Ramakers, C., J. M. Ruijter, R. H. Lekanne Deprez, A. F. M. Moorman (2003). Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. Neuroscience Letters **339**, 62-66.

Rutledge, R. G. (2004). Sigmoidal curve-fitting redefines quantitative real-time PCR with the prospective of developing automated high-throughput applications. Nucleic Acids Res. **32**, e178.

Saiki, R., S. Scharf, F. Faloona, K. Mullis, G. Horn, H. Erlich (1985). Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science **230**, 1350-54.

Stone, E., J. Goldes, M. Garlick (2007). A multi-stage model for quantitative PCR. In review.

Sun, F., D. Galas, M. Waterman (1996). A mathematical analysis of in vitro molecular section-amplification. J. Theor. Bio. **258**, 650-660.

Valasek, M. A. and J. J. Repa (2005). The power of real-time PCR. Advan. Physiol. Educ. **29**, 151-159.

Velikanov, M. and R. Kapral (1999). Polymerase chain reaction: A markov process approach. J. Theor. Bio. **201**, 239-249.

Weiss, G. and A. von Haeseler (1995). Modeling the polymerase chain reaction. J. Comput. Biol. **2**, 49-61.