Modeling DNA Amplification Technology for Process Optimization.

- **Mentor:** Dr. Emily Stone, University of Montana.

- **Group**
  - Neil Olver, McGill University
  - Jon Collis, RPI
  - Yevgeniy Frenkel, RPI
  - Derek Moulton, University of Delaware
  - Ryan Haskett, Duke University
  - John Evans, RPI
  - Fritzner Soliman, RPI
  - Amal Aafif, Drexel University
  - Dorjsuren Badamдорж, University of Cincinnati.

**Abstract**

The objective of this problem is to study PCR reactions in the Ruggedized Advances Pathogen Identification Device (RAPID) Cycler in order to model the reactions and optimize the process of DNA amplification. The RAPID Cycler is a devise designed by Idaho Technologies (IT) to identify DNA sequences using the Polymerase Chain Reaction (PCR) method. This method consists in three basic stages: melting, annealing and extension. In these steps different reactions between the DNA sequences and the reactives take place. These reactions are modeled as sets of ODEs with different reaction rates. The goals in this project are, on one hand, to solve numerically the ODEs in order to fit the reaction rates with the data from IT. On the other hand, the goal is to optimize the time required in each stage of the PCR method, searching for an efficient process. The group modeled the reactions using two approaches. In the temperature varying model, the reaction rates were treated as functions of the temperature and all reactions occurred simultaneously. In the reduced model, each reaction occurs separately. An analytic solution of the equations was found for this model. Once the models were formulated, the parameters (reaction rates) were fit using the IT data. To optimize the process, three different approaches were taken. In the fixed time optimization, the idea is to fix the times for each stage in order to minimize the time of the
complete process. The stage time optimization allows different times on each stage in order to optimize the efficiency of the process. In the global stage time optimization, each stage is individually optimized. Finally, the Taq-on/Taq-off model was studied.

1 Problem Description.

In this project, the objective is to study PCR reactions in the Ruggedized Advances Pathogen Identification Device (RAPID) Cycler in order to model the reactions and optimize the process of DNA amplification. PCR stands for Polymerase Chain Reaction. It is an in vitro method for the enzymatic synthesis of specific DNA sequences invented in 1985 by Kary Mullis, who won the Nobel Prize for this work. Its principal applications lie in the field of molecular biology in both basic research and applications in medicine (diagnosis) or pathogen detection and DNA sequencing (e.g. fingerprinting). In DNA recognition, the goal is to determine the presence of a certain agent (a pathogen agent, for instance, such as anthrax, etc.) in a sample. This is done by amplifying specific sequences of DNA that are characteristic and unique of the target’s DNA (composed of about 200-500 base pairs). A key element in this process is the utilization of a Taq Polymerase, which is a high temperature stable enzyme which allows the amplification of sequences, given that sequences other than the target will not form “accidentally” by themselves in a high temperature environment.

The RAPID Cycler is a device designed by Idaho Technologies (IT) to identify DNA sequences using the PCR method. This instrument was engineered to match the speed of biochemical reactions given its rapid temperature cycling system. This provides nearly instantaneous temperature transactions, ensures uniform temperature and rapid heat exchange within the sample.

The basic steps on the PCR reaction are

1. Melting, the sample is heated to 94°C to denature DNA strands.
2. Annealing, the sample is cooled to 60°C, at this stage primer molecules bind on to each end of a target sequence.
3. Extension, heat the sample again to 70°C, the Taq polymerase is used and single nucleotides are added to the single strand sequence.

These steps are repeated 30–40 times to obtain 80% yield (or 80% of primer molecules).

In the Quantitative Real Time PCR, fluorescent probes are used to detect completed strands. TaqMan probes are labeled on the 5 minute end with a reporter dye and on the 3 minute end with a quencher. During PCR, the fluorogenic probe hybridizes between the forward and reverse primers. As Taq DNA polymerase extends, the TaqMan probe is cleaved, separating the reporter from the quencher and generates a fluorescent signal proportional to the number of amplicons produced. The Light Typer is a device designed to detect SPN (single nucleotide polymorphisms), which detects completed strands through
fluorescing probe molecules. In Figure 2, we observe the amplification curves for the Light Cycler in the quantitative PCR. The goal is to determine the initial amount of DNA target present in the sample.

The simplest model to quantify the initial amount of target DNA sequences in the sample is the “simple doubling,” where the copy number of sequences $C(n)$ at the cycle $n$ is given in terms of the initial amount of copies $C(0)$ as

$$C(n) = 2^nC(0).$$

In Figure 3, we find a “standard curve” of the time to cross a set threshold fluorescence value as a function of the initial copy number using this model (“Famous Model”).
Several problems arise in the current PCR reaction modeling, such as the identification of the exponential region, locating a good fluorescence threshold for all curves, errors in the initial copy number which are amplified and a saturation in the process that is clearly observed (which the model does not predict). This leads to the conclusion that the process is not exactly logistic, especially for low copy number curves. In previous work (GIMMC, June 2003), students tried different methods for data fitting the logistic model, some results are shown if Figure 4.

In order to improve the model to optimize the reaction design, the conditions need to be finely tuned for any new sequence to be amplified. This can take weeks in the lab. A good model could reduce this time by narrowing regions of parameter space to be explored in the laboratory, with the purpose of optimizing both the yield and the specificity of the reaction. We can write down the reaction that takes place in the PCR method:

**Reaction Phase**

| Dissociation | ds DNA $\xrightarrow{d}$ 2 ss DNA |
| Priming      | P + ss DNA $\xrightarrow{k_1}$ (ss DNA)$' + b$ |
| Extension    | R+ (ss DNA)$'$ $\xrightarrow{k_2}$ ds DNA + P. |

Let $A = [\text{ds DNA}]$ be the concentration of double stranded DNA, $B = [\text{ss DNA}]$ be the concentration of single stranded DNA, $B' = [(\text{ss DNA})']$ be the concentration of single stranded DNA once the annealing process is concluded, $P = [P]$ be the concentration of the primer and $R = [R]$ be the concentration of the resource. Then, these reactions may be modeled as a set of coupled ODEs with temperature dependent reaction rates. According to the stage in the PCR cycle (given by the temperature), different reactions “turn on and
Figure 4: a) Fitting the logistic model: least squares and nonlinear optimization, logistic map: $y_{n+1} = y_n + ry_n(k-y_n)$.  

b) Fitting the Taq model (modified efficiency).
off”. We may write,

\[
\begin{align*}
\frac{dA}{dt} &= -k_1 A \\
\frac{dB}{dt} &= 2k_1 A \\
\frac{dP}{dt} &= -k_2 BP \\
\frac{dB}{dt} &= -k_2 BP \\
\frac{dB'}{dt} &= k_2 BP \\
\frac{dR}{dt} &= -k_3 RB' \\
\frac{dB'}{dt} &= -k_3 RB' \\
\frac{dA}{dt} &= k_3 RB'.
\end{align*}
\]

| $k_1 > 0, k_2, k_3 \approx 0$ | Dissociation ($T = 94^\circ C$) |
| $k_2 > 0, k_1, k_3 \approx 0$ | Priming ($T = 60^\circ C$) |
| $k_3 > 0, k_1, k_2 \approx 0$ | Extension ($T = 70^\circ C$) |

In Figure 5 we see a sketch of the temperature during one cycle of PCR. The goal of the optimization process is to minimize the time required for the incorporation of primers and to maximize the time required for the yielding process (divided in $n$ cycles during $t$ seconds each), using a two stage approximation model parametrized from IT data.

The goals to be achieved in this project consist of modeling the reactions as ODEs. These equations must be solved numerically in Matlab in order to fit the parameters (reaction rates) with the data from IT. “Brute force” optimization may be used for the sampling space of varying stage I and stage II times (the annealing and extension stages, respectively). Compute the time to completion, complemented or performed using a Matlab optimization code, to solve for the times that give the least time to completion.
2 Analysis.

2.1 Modeling Equations.

To model the PCR process, the following reactions are considered

<table>
<thead>
<tr>
<th>Reaction</th>
<th>D</th>
<th>k₁</th>
<th>2S</th>
<th>S'</th>
<th>k₄</th>
<th>S+P</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Melt”</td>
<td></td>
<td>k₁</td>
<td>2S</td>
<td>S'</td>
<td>k₄</td>
<td>S+P</td>
</tr>
<tr>
<td>“Anneal”</td>
<td>S+P</td>
<td>k₂</td>
<td>S'</td>
<td>S'</td>
<td>k₃</td>
<td>D</td>
</tr>
<tr>
<td>“Extend”</td>
<td>S+R</td>
<td>k₃</td>
<td>S'</td>
<td>S'</td>
<td>k₃</td>
<td>D</td>
</tr>
</tbody>
</table>

where D stands for doubled stranded DNA, S for single stranded DNA, P for the primers, R for the nucleotides (DNA base pairs), S' for the primed single stranded DNA and kᵢ for the reaction rates. Denoting with lowercase letters the concentrations of each reactive, we may write the following differential equations for each stage:

\[
\frac{dd}{dt} = -k₁d, \quad \frac{ds}{dt} = 2k₁d + k₄s', \quad \frac{ds'}{dt} = -k₄s', \quad \frac{dp}{dt} = k₄s'.
\]

for the melting stage, and

\[
\frac{dp}{dt} = -k₂sp, \quad \frac{ds}{dt} = -k₂sp, \quad \frac{ds'}{dt} = k₂sp,
\]

for the annealing stage, with initial conditions

\[
s_{I}^{I}(0) = 2d_{i-1}^{II}(t_{II}) + s_{i-1}^{II}(t_{II}), \quad p_{I}^{I}(0) = p_{i-1}^{I}(t_{I}) + s_{i-1}^{II}(t_{II}), \quad s_{I}^{I}(0) = 0,
\]

where i stands for the cycle number, I and II are the annealing and extension stages, respectively and t_I and t_{II} are the times at the end of these stages. For the extension stage we consider the differential equations

\[
\frac{ds'}{dt} = -k₃rs, \quad \frac{dr}{dt} = -k₃rs, \quad \frac{dd}{dt} = k₃rs,
\]
with initial conditions,
\[ d_i^{II}(0) = 0, \quad r_i^{II}(0) = r_i^{II}(t_{II}), \quad s_i^{\prime\prime}(0) = s_i^{II}(t_i). \]

2.1.1 Two Models for the Reactions.

Two models were proposed to describe the reaction. In the temperature varying model, the rate constants for reactions are treated as functions of temperature. All reactions occur simultaneously and a normal distribution is assumed for each rate constant with mean given by PCR protocol. The advantage of this model is that it can incorporate different temperature curves over time. Its disadvantage is the slow computation time and the fact that no analytic solution can be found. In the reduced model, each stage occurs separately. It is assumed that the melting stage provides perfect dissociation (i.e. melting stage provides initial conditions for annealing), the annealing reaction rate is constant and in the extension stage the reaction rate decreases by 2% each cycle (due to decrease in Taq). For this model, the analytical solution is given by
\[
\begin{align*}
p &= (p_0 - s_0) \left[ 1 + \frac{p_0}{s_0 e^{(p_0-s_0)k_2 t}} - 1 \right], \\
s &= \frac{p_0 - s_0}{s_0 e^{(p_0-s_0)k_2 t} - p_0}, \\
s' &= s_0 - \frac{p_0 - s_0}{s_0 e^{(p_0-s_0)k_2 t} - p_0}.
\end{align*}
\]

Similar solutions are found for \( r, s' \) and \( d \) in stage II. A numerical comparison between the two models can be done by considering the temperature distribution as showed in Figure 5. The comparison of the two models is presented in Figure 6, showing that the reduced model’s curve lags behind the temperature varying model.

2.2 Parameter Fitting.

The next step in the analysis was to fit the parameters using the IT data. This was done by using a least squares approach. There are two separate types of parameters to fit. In the first type, we consider \( k_2 > k_3 \), for the second type, we consider \( k_3 > k_2 \). The curve fitting of the parameters for the first type is shown in Figure 7. It is then found that
\[
k_2 = 0.0084, \quad k_3 = 0.0029, \quad d_0 = 0.0032, \quad P_0 = 8.1, \quad r_0 = 20.
\]

For the second type, Figure 8 shows the curve that fits the parameters. The following are the fitted parameters,
\[
k_2 = 0.30, \quad k_3 = 1.45, \quad d_0 = 10^{-4}, \quad P_0 = 0.1, \quad r_0 = 1.
\]
2.3 Optimization.

The next step was to optimize the process. In fixed time optimization the goal is to optimize fixed $t_I$ and $t_{II}$ for the two data fits. The minimization criterion is to minimize the total time until 80% of the primer is exhausted. The basic idea is that since the minimization surface is rather “bumpy,” we first find candidates for a minimum by a sampling process and then run Matlab’s fminsearch on these candidates. For the first fit type, the optimization surface is shown in Figure 9, giving $t_I = 22.6587$ seconds and $t_{II} = 22.6916$ seconds, for a total time of 852 seconds for the whole process.

For the second fit type the optimization surface is shown in Figure 10. We found that $t_I = 10.8622$ seconds and $t_{II} = 2.5419$ seconds, for a total time of 323 seconds for the whole process. In the comparison between the two fitting types, it was concluded that the optimization for the fit type one results in a better approach since it is more similar to the actual process.

In stage time optimization the reduced model was used. Now, the annealing and extension times are allowed to vary over different cycles in order to optimize the efficiency in each individual cycle according to

$$\text{Efficiency} = \frac{\text{primer used}}{t_I + t_{II} + t_{melt}}.$$  

The efficiency surface is shown in Figure 11. This surface varies according to the initial conditions which were calculated at each cycle. The time protocol for this optimization approach is shown in Figure 12, giving a total time of 796 seconds for the entire process.

In the global stage time optimization the reduced model was used and each stage time was individually optimized. This results in $2 \times \text{(number of cycles)}$ variables to optimize. To perform this optimization model a simulated annealing was used, leading to a N-dimensional optimization problem with random steps. The idea is to jump between local minima in search.
Figure 7: Parameters for Fit type 1.

Figure 8: Parameters for Fit type 2.

Figure 9: Optimization surface for fixed time optimization (fit type 1).
Figure 10: Optimization surface for fixed time optimization (fit type 2).

Figure 11: Optimization surface for the stage time optimization.
of a global minimum. Figure 13 shows the optimal stage times for each cycle according to the amount of primary used. At the end of the graph we observe that the optimization method gives a zero time for stage I. This is because the primary is exhausted and then a zero time for stage I at this point is required. Nonetheless, a 10% gain in efficiency is found using this optimization scheme. In Figure 14, the optimal annealing time by cycle is plotted as a function of the number of cycles, showing an exponential growth.

2.3.1 Taq-on/Taq-off.

Finally, a *Taq-on/Taq-off* model was studied. Here, a reaction with Taq (Q) polymerase is accounted for in the extension stage forming a complex (C).

\[
S' + Q \xrightarrow{k_3} C \xrightarrow{k_4} D + Q
\]

This model leads to a set of coupled ODEs which cannot be solved analytically. This model is not suitable for a large number of cycles.

3 Conclusions.

During the week, two models were proposed to describe the reactions that take place in each cycle of the RAPID cycler. The temperature varying model incorporates different temperature curves over time, but the computation time is poor and there is no analytic solution. For the reduced model, an analytical solution could be found. A numerical comparison between these two models was done, showing that the reduced model’s curve lags behind the curve of the temperature varying model. A least squares approach was used to fit the parameters for the \( k_2 > k_3 \) and \( k_3 > k_2 \) cases. Three different optimization criteria were used. In the fixed time optimization, the annealing and extension times are fixed and the goal is to minimize the total process time until 80% of the primer is exhausted. It was found that the \( k_2 > k_3 \) case results in a better approach since it is more similar to the actual process. In the stage time optimization, the reduced model is used and the criterion was to optimize the efficiency in each individual cycle. This efficiency was chosen to be the ratio of the primer used and the three stage times. For this criterion, the total time of the process was found to be 796 seconds. The global stage time optimization used the reduced model and each stage time was individually optimized. Using this criterion, a 10% gain in efficiency was found. Finally, the Taq-on/Taq-off model was studied.
Figure 12: Time protocol for the stage time optimization.

Figure 13: Optimal stage time for each cycle in the global stage time optimization.

Figure 14: Optimal annealing time for each cycle in the global stage time optimization.