

Mathematical Model of Real-Time PCR Kinetics

Jana L. Gevertz,¹ Stanley M. Dunn,¹ Charles M. Roth^{1,2}

¹Department of Biomedical Engineering, Rutgers University, Piscataway, New Jersey

²Department of Chemical & Biochemical Engineering, Rutgers University, 98 Brett Road, Piscataway, New Jersey 08854, telephone: 732-445-4109; fax: 732-445-2581; e-mail: cmroth@rci.rutgers.edu

Received 17 February 2005; accepted 16 May 2005

DOI: 10.1002/bit.20617

Abstract: Several real-time PCR (rtPCR) quantification techniques are currently used to determine the expression levels of individual genes from rtPCR data in the form of fluorescence intensities. In most of these quantification techniques, it is assumed that the efficiency of rtPCR is constant. Our analysis of rtPCR data shows, however, that even during the exponential phase of rtPCR, the efficiency of the reaction is *not* constant, but is instead a function of cycle number. In order to understand better the mechanisms belying this behavior, we have developed a mathematical model of the annealing and extension phases of the PCR process. Using the model, we can simulate the PCR process over a series of reaction cycles. The model thus allows us to predict the efficiency of rtPCR at any cycle number, given a set of initial conditions and parameter values, which can mostly be estimated from biophysical data. The model predicts a precipitous decrease in cycle efficiency when the product concentration reaches a sufficient level for template–template re-annealing to compete with primer–template annealing; this behavior is consistent with available experimental data. The quantitative understanding of rtPCR provided by this model can allow us to develop more accurate methods to quantify gene expression levels from rtPCR data. © 2005 Wiley Periodicals, Inc.

Keywords: gene expression; mathematical modeling; PCR efficiency

INTRODUCTION

Real-time PCR (rtPCR) is a tool that has gained widespread application in recent years for measuring the expression levels of individual genes. While techniques, such as Northern blotting, have existed for decades to perform this task, rtPCR has the ability to measure small changes in gene expression and to quantify rare transcripts—both of which are extremely difficult using standard Northern blotting techniques (Pfaffl, 2001; Roth, 2002). Furthermore, due to its

quantitative nature, rtPCR has become the gold standard in validating DNA microarray results (Stagliano et al., 2003).

While rtPCR is considered to be quantitative, it is still necessary to relate experimental data, in the form of fluorescence intensities, to initial mRNA levels via a mathematical model. Equations have been developed for either absolute or relative quantification of rtPCR data, which require the comparison of samples to purified standards or housekeeping genes, respectively (Goidin et al., 2001; Rutledge and Cote, 2003; Schmittgen and Zakrajsek, 2000). Early approaches, still adopted by many users, quantify relative expression based on the change in threshold cycle, $2^{-\Delta\Delta C_T}$ (Livak and Schmittgen, 2001), where C_T is the threshold cycle number (where the fluorescence reaches an arbitrary level within the exponential growth region) and the factor two implies a perfect efficiency at each cycle used for quantitation. It is recognized that PCR is not perfectly efficient, and various analyses have been proposed for determining a value of efficiency based either on dilution series or directly from the log(fluorescence) versus cycle number data (Liu and Saint, 2002; Ramakers et al., 2003; Tichopad et al., 2003). Some of the experimental factors limiting the efficiency of PCR include the initial concentrations of starting material (Taq polymerase, primers, deoxyribonucleotides, and template DNA), the degradation of Taq polymerase, PCR product reannealing and primer-dimer accumulation (Roth, 2002). Since the amounts of each of the biochemical species change from cycle-to-cycle, PCR efficiency is likely to vary systematically during the course of a run. For this reason, we have analyzed the efficiency of each PCR cycle using experimental data. A previous study calculated the efficiency for one set of data and found that it is indeed not constant before the plateau, but is instead a function of cycle number (Peirson et al., 2003).

This observation has prompted us to develop a mathematical model of PCR efficiency with respect to cycle number. In order to develop a theoretical model that can predict the efficiency of PCR, we have represented the reactions occurring at each PCR stage using equilibrium and kinetic equations. Specifically, we consider the steps of PCR—DNA

Correspondence to: Charles M. Roth

Jana L. Gevertz's present address is Program in Applied and Computational Mathematics, Princeton University, Princeton, New Jersey.

Contract grant sponsor: NSF CAREER Award

Contract grant number: BES-0238617

denaturation, primer annealing, and DNA extension. In particular, we focus mostly on the annealing step and address the extent to which equilibrium is reached during the annealing phase of PCR. From the thermodynamics and kinetics of the PCR reactions, we can simulate the PCR process and calculate the efficiency as a function of cycle number. We propose that such a model will be useful in establishing guidelines for the optimization of PCR protocols as well as selecting among the various quantitation techniques available to estimate expression levels from fluorescence output.

METHODS

Analysis of rtPCR Data

Real-time PCR exploits the PCR process in order to quantify gene expression in samples from cells or tissues. After extraction of sample RNA, the first step in rtPCR is the reverse transcription of the target mRNA. In theory, the resulting cDNA concentration should be equal to the concentration of target mRNA. In practice, however, the resulting concentration of cDNA, $[\text{DNA}]_0$, is less than the initial concentration of mRNA, $[\text{RNA}]_0$. That is,

$$[\text{DNA}]_0 = [\text{RNA}]_0 * \varepsilon_{\text{RT}} \quad (1)$$

where ε_{RT} is the efficiency of the reverse transcription step. It has been reported that ε_{RT} may vary between 5% and 90% (Freeman et al., 1999); however, with application of a consistent protocol the error in a particular experimental system should be considerably less.

For the subsequent PCR amplification, however, small changes in efficiency will propagate. In the idealized case, after each PCR cycle the amount of target DNA present will double, and the total concentration of DNA present after any cycle n would be given by,

$$[\text{DNA}]_n = [\text{DNA}]_0 \times 2^n \quad (2)$$

where $[\text{DNA}]_0$ is the initial concentration of DNA. This theoretical concentration is rarely achieved, however, at all but the earliest cycles of PCR (Roth, 2002). As a result, many analyses of rtPCR account for the imperfect efficiency, ε , of the PCR reaction, as,

$$[\text{DNA}]_n = [\text{DNA}]_0 * (1 + \varepsilon)^n \quad (3)$$

The efficiency is assumed to be constant for each PCR cycle, but it is applied only in the region before a plateau is reached (Liu and Saint, 2002; Tichopad et al., 2003).

In rtPCR, the accumulation of PCR products is monitored at the end of each cycle by fluorescence detection. Ideally, after subtracting out background fluorescence, the level of fluorescence at the end of a PCR cycle (F_n) is directly proportional to the concentration of product at that cycle, that is,

$$F_n = \alpha * [\text{DNA}]_n = \alpha * [\text{DNA}]_0 * (1 + \varepsilon)^n \quad (4)$$

where α is a proportionality constant. Use of this method to accurately determine $[\text{DNA}]_0$ (and hence gene expression levels) from fluorescence intensities requires a value of the constant ε . This can be determined by a fit to the fluorescence versus cycle number data, usually in a log-linear form,

$$\log(F_n) = \log[\alpha[\text{DNA}]_0] + n \log(1 + \varepsilon) \quad (5)$$

This assumes an efficiency that is not a function of cycle number.

Relating DNA Levels and Fluorescence Levels

In general, each cycle of PCR can have a unique efficiency value, which we will define as $\varepsilon_{\text{PCR}}^{(n)}$. The efficiency can be evaluated directly from background subtracted fluorescence vs. cycle number data, that is,

$$\varepsilon_{\text{PCR}}^{(n)} = \frac{F_n}{F_{n-1}} - 1 \quad (6)$$

If the efficiency of PCR is not constant, we can account for the varying efficiencies explicitly at each cycle,

$$F_n = \alpha[\text{DNA}]_0 \prod_{i=1}^n \left(1 + \varepsilon_{\text{PCR}}^{(i)}\right) \quad (7)$$

where $\varepsilon_{\text{PCR}}^{(i)}$ is the efficiency of the i th PCR cycle.

Experimental Methods

Several rtPCR experiments were performed using 18S cDNA at varying starting concentrations as described previously (Roth, 2002). After subtracting off background fluorescence, Equation (6) was used to determine $\varepsilon_{\text{PCR}}^{(n)}$ for each PCR cycle performed.

Model Development

In an effort to determine the efficiency of PCR at each cycle n , the process of PCR was broken down into three steps: double-stranded (ds) DNA denaturation, primer annealing, and DNA extension (Fig. 1). By determining the efficiency of each stage of PCR, we will, in turn, have a model for $\varepsilon_{\text{PCR}}^{(n)}$.

Step 1: DNA Denaturation

Experimentation has shown that when dsDNA is subjected to temperatures ranging from 94–97°C (the typical range for the denaturing temperature) for five or more seconds, it is completely denatured. Thus, we assume that the efficiency of denaturation at each cycle n is unity,

$$\varepsilon_{\text{den}}^{(n)} = 1 \quad (8)$$

Step 2: Primer Annealing

After the dsDNA is denatured, two complementary template strands of DNA, which we will denote T_1 and T_2 , are

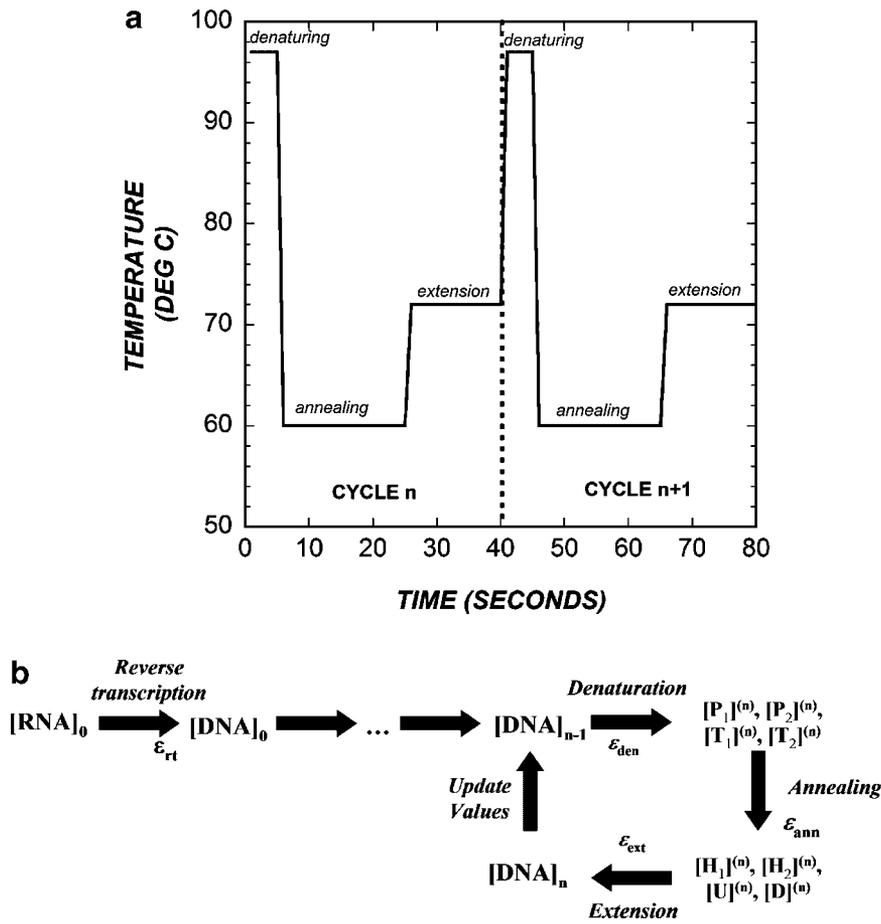


Figure 1. a: Temperature cycling in PCR; (b) flow of species in model of rtPCR.

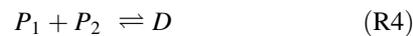
produced. Once the temperature is cooled to the annealing temperature, a strand-specific primer, P_1 will anneal to T_1 to form a hybrid (H_1), and likewise P_2 and T_2 will anneal to form a hybrid (H_2). These two reactions can be represented by the chemical equations,



Simultaneously, other reactions can occur during the annealing stage of PCR. The template strands, T_1 and T_2 , are complementary and can re-anneal upon contact, forming a template hybrid (U),



Furthermore, depending on primer design, primer-dimers (D) can form as well.



Hence, at least four reactions can occur simultaneously during the primer annealing stage. These reactions were modeled using both equilibrium and kinetic descriptions of the reactions.

Step 2A: Equilibrium Model

In order to track the total concentration of all products performing active roles during the annealing stage of PCR, a mass balance is performed on each of the primers and templates. This procedure results in the following four equations:

$$[P_1]_T = [P_1] + [H_1] + [D] \quad (9)$$

$$[P_2]_T = [P_2] + [H_2] + [D] \quad (10)$$

$$[T_1]_T = [T_1] + [H_1] + [U] \quad (11)$$

$$[T_2]_T = [T_2] + [H_2] + [U] \quad (12)$$

where $[X]_T$ denotes the total concentration of species X .

Here, each of the reactions (R1)–(R4) is assumed to proceed to equilibrium. For each reaction, the ratio of reactant to product concentrations is fixed and equal to the equilibrium constant (written for dissociation of each hybrid),

$$K_{H_1} = \frac{[P_1][T_1]}{[H_1]} \quad (13)$$

$$K_{H_2} = \frac{[P_2][T_2]}{[H_2]} \quad (14)$$

$$K_U = \frac{[T_1][T_2]}{[U]} \quad (15)$$

$$K_D = \frac{[P_1][P_2]}{[D]} \quad (16)$$

However, the value of K_U would be extremely low (on the order of 10^{-90} M at 55°C for a template of length 150 bp), due to the accumulation of free energy along the length of the template–template duplex. As a result, incorporation of template–template binding in an equilibrium model would dominate over primer–template binding and lead to the false conclusion that only template–template reannealing occurs and PCR does not proceed. Therefore, we removed Equation (15) from the equilibrium model and only considered it in the kinetic model.

Step 2B: Kinetic Model

If the reactions (R1)–(R4) do not proceed to equilibrium, we must follow their progression in time. Each equilibrium constant in Equations (13)–(16) is the ratio of kinetic dissociation and association rate constants, for example, for K_{H_1} ,

$$K_{H_1} = \frac{k_{dH_1}}{k_{aH_1}} \quad (17)$$

where the lower case k denotes a rate rather than equilibrium constant and the d and a in the subscripts denote dissociation and association, respectively.

By applying mass action kinetic balances to each species, the reactions are described by a nonlinear system of eight differential equations,

$$\frac{d[P_1]}{dt} = -k_{aH_1}[P_1][T_1] + k_{dH_1}[H_1] - k_{aD}[P_1][P_2] + k_{dD}[D] \quad (18)$$

$$\frac{d[P_2]}{dt} = -k_{aH_2}[P_2][T_2] + k_{dH_2}[H_2] - k_{aD}[P_1][P_2] + k_{dD}[D] \quad (19)$$

$$\frac{d[T_1]}{dt} = -k_{aH_1}[P_1][T_1] + k_{dH_1}[H_1] - k_{aU}[T_1][T_2] + k_{dU}[U] \quad (20)$$

$$\frac{d[T_2]}{dt} = -k_{aH_2}[P_2][T_2] + k_{dH_2}[H_2] - k_{aU}[T_1][T_2] + k_{dU}[U] \quad (21)$$

$$\frac{d[H_1]}{dt} = k_{aH_1}[P_1][T_1] - k_{dH_1}[H_1] \quad (22)$$

$$\frac{d[H_2]}{dt} = k_{aH_2}[P_2][T_2] - k_{dH_2}[H_2] \quad (23)$$

$$\frac{d[U]}{dt} = k_{aU}[T_1][T_2] - k_{dU}[U] \quad (24)$$

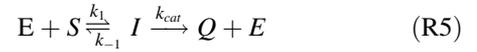
$$\frac{d[D]}{dt} = k_{aD}[P_1][P_2] - k_{dD}[D] \quad (25)$$

We used a value of $k_{dU}=0$ to indicate the essentially irreversible nature of template–template reannealing.

For the first PCR cycle, the initial condition on each of the reactants and products is imposed by the amount of each primer added to the PCR reaction and the amount of template in the sample after reverse transcription. For subsequent cycles, these amounts are updated as described below.

Step 3: DNA Extension/Synthesis

In the extension reaction, the Taq polymerase enzyme (E) binds to primer–template hybrids (denoted S in this stage) and is responsible for joining nucleotides together to synthesize product dsDNA (Q). This process is represented by the following chemical reaction scheme:



where I is an intermediary complex of S and E in which extension has not yet occurred.

Clearly, the concentration of active enzyme plays an important role in this reaction. If the enzyme concentration falls below some critical value, a decrease in $\epsilon_{\text{ext}}^{(n)}$ will be observed. Although Taq polymerase is used because of its thermostability, some enzyme deactivation may occur as rtPCR is carried out, because denaturation of template is conducted at such high temperatures ($94\text{--}97^\circ\text{C}$). We assume that Taq polymerase is deactivated according to first order kinetics,

$$[E]^{(n+1)} = [E]^{(n)} \exp(-k_{\text{deg}} t_{\text{den}}) \quad (26)$$

where k_{deg} is the degradation constant of the enzyme ($k_{\text{deg}} > 0$) and t_{den} is the amount of time (in seconds) that the enzyme is exposed to the denaturation temperature in any cycle. For the cycle time in our laboratory protocol and the statistics on Taq DNA Polymerase (Sambrook and Russell, 2001; Qiagen, 2002), only 0.1% of the enzyme activity is lost at each step,

$$\frac{[E]^{(n)}}{[E]^{(0)}} = (0.9990)^n \quad (27)$$

Reaction (R5) was analyzed using Michaelis–Menten kinetics with two key modifications. First, the amount of active enzyme (decreasing slightly at each cycle) was accounted for as described by Equation (27). Second, a

time-dependent version of the Michaelis–Menten formalism was used to account for the fact that, in PCR, the substrate is in lower concentration than the enzyme. The substrate concentration is thus described by,

$$-\frac{d[S]}{dt} = v_0 = \frac{V_{\max}[S]}{K_m + [S]} \quad (28)$$

where $V_{\max} = k_{\text{cat}}[E]^{(n)}$, $K_m = \frac{k_{-1} + k_{\text{cat}}}{k_1}$, and the substrate concentration $[S]$ includes both primer-template hybrids (H_1 and H_2). That is,

$$[S]_{\text{T}}^{(n)} = [H_1]^{(n)} + [H_2]^{(n)} \quad (29)$$

The solution to the differential Equation (28) is,

$$\ln\left(\frac{[S]_{\text{T}}}{[S]}\right) + \frac{[S] - [S]_{\text{T}}}{K_m} = \frac{V_{\max}t}{K_m}. \quad (30)$$

The root of this implicit expression for $[S]$ on $\{0, [S]_{\text{T}}\}$ was found numerically in Maple.

Model Implementation and Calculation of Efficiencies

The efficiency of the n th annealing stage ($\varepsilon_{\text{ann}}^{(n)}$) was calculated by comparing the amount of hybrids after the n th annealing stage to the total amount of template present throughout the n th annealing stage,

$$\varepsilon_{\text{ann}}^{(n)} = \frac{[H_1]^{(n)} + [H_2]^{(n)}}{[T_1]_{\text{T}}^{(n)} + [T_2]_{\text{T}}^{(n)}} \quad (31)$$

where $[H_1]^{(n)}$ and $[H_2]^{(n)}$ were calculated by solving the governing thermodynamic (Eqs. 9–16) or kinetic (Eqs. 18–25) relations. Although Equations (9)–(16) can be solved analytically, the resulting expression is very complex, leading to significant numerical error accumulation in its evaluation. For this reason, it proved more useful to solve the system of equations numerically for $[H_1]$ and $[H_2]$; this was done using Maple. In order to solve the system of kinetic (differential) equations, Maple’s “dsolve” function, which uses the Fehlberg fourth-fifth order Runge–Kutta method with degree four interpolant, was employed.

The efficiency of DNA extension at cycle n , $\varepsilon_{\text{ext}}^{(n)}$, was calculated by computing the fraction of substrate converted during the extension phase of the cycle,

$$\varepsilon_{\text{ext}}^{(n)} = \frac{[S]_{\text{T}}^{(n)} - [S]_{\text{ext}}^{(n)}}{[S]_{\text{T}}^{(n)}}. \quad (32)$$

Given the efficiencies of annealing and extension, and assuming that denaturation is perfectly efficient, the overall efficiency for cycle n is simply,

$$\varepsilon_{\text{PCR}}^{(n)} = \varepsilon_{\text{ann}}^{(n)} \varepsilon_{\text{ext}}^{(n)}. \quad (33)$$

After finding the solution of the model equations at a particular cycle n , the values of primer and template concentrations are updated before proceeding to cycle $n+1$. The concentration of each primer at cycle $n+1$ is determined by the total concentration at cycle n , minus the primer incorporated into new templates at cycle n ,

$$[P_1]_{\text{T}}^{(n+1)} = [P_1]_{\text{T}}^{(n)} - \varepsilon_{\text{ext}}^{(n)}[H_1]^{(n)} \quad (34)$$

$$[P_2]_{\text{T}}^{(n+1)} = [P_2]_{\text{T}}^{(n)} - \varepsilon_{\text{ext}}^{(n)}[H_2]^{(n)} \quad (35)$$

where $\varepsilon_{\text{ext}}^{(n)}$ is the efficiency of the n th DNA extension process. Observe that primer-dimers are not subtracted off, as they are dissociated after the denaturing stage of PCR (right before the annealing stage at cycle $n+1$)—resulting in no net loss of primers. The template concentrations are updated as,

$$[T_1]_{\text{T}}^{(n+1)} = [T_1]_{\text{T}}^{(n)} + \varepsilon_{\text{ext}}^{(n)}[H_2]^{(n)} \quad (36)$$

$$[T_2]_{\text{T}}^{(n+1)} = [T_2]_{\text{T}}^{(n)} + \varepsilon_{\text{ext}}^{(n)}[H_1]^{(n)}. \quad (37)$$

Note that the concentration of T_1 is increased by the amount of H_2 (and not H_1), because extending an H_2 hybrid results in the net gain of a T_1 but no net gain of a T_2 .

Finally, we assume that exposure to the denaturing temperature for 5 s denatures all primer-template hybrids, primer-dimers and template hybrids. Therefore, for each cycle n ,

$$[H_1]_{\text{T}}^{(n)} = [H_2]_{\text{T}}^{(n)} = [U]_{\text{T}}^{(n)} = [D]_{\text{T}}^{(n)} = 0. \quad (38)$$

Parameter Determination

Table I gives a complete listing of all parameters and values used in the model. These are taken from standard laboratory protocols and, for thermodynamic and kinetic parameters, values estimated for the 18S cDNA system (Roth, 2002). $[P_1]_{\text{T}}^{(1)}$ and $[P_2]_{\text{T}}^{(1)}$ are simply the initial concentration of primers put into the rtPCR system. The values of $[T_1]_{\text{T}}^{(1)}$ and $[T_2]_{\text{T}}^{(1)}$ represent the expected range of starting template concentrations, which may vary considerably in samples analyzed by rtPCR. The equilibrium constants (K_{H_1} , K_{H_2} , K_U , K_D) were computed for 18S primers based on nearest-neighbor thermodynamics (Tinoco et al., 2001). Since K_U was negligibly small, the template–template reaction was eliminated in the equilibrium formulation. The values of the association rate constants (k_{aH_1} , k_{aH_2} , k_{aU} , k_{aD}) were based on the experimental literature for nucleic-acid association (Plum et al., 1999), and the values of the dissociation rate constants (k_{dH_1} , k_{dH_2} , k_{dU} , k_{dD}) were determined by the relationship between the equilibrium constant of the reaction, K_{eq} , and the association rate constant, k_a (Eq. (17)).

Table I. Parameter values.

Parameter	Values(s)	Source
$[P_1]_T^{(1)}$	10^{-6} M	PCR protocol
$[P_2]_T^{(1)}$	10^{-6} M	PCR protocol
$[T_1]_T^{(1)}$	10^{-14} M $\leq [T_1]_T^{(1)} \leq 10^{-10}$ M	Estimated based on practice
$[T_2]_T^{(1)}$	10^{-14} M $\leq [T_2]_T^{(1)} \leq 10^{-10}$ M	Estimated based on practice
K_{H_1}	5.5531×10^{-13} M	Calculated ^a
K_{H_2}	8.1493×10^{-11} M	Calculated ^a
K_U	≈ 0	Calculated ^a
K_D	10^{-2} M	Calculated ^a
k_{aH_1}	10^6 M ⁻¹ s ⁻¹	Plum et al. (1999)
k_{aH_2}	10^6 M ⁻¹ s ⁻¹	Plum et al. (1999)
k_{aU}	10^6 M ⁻¹ s ⁻¹	Plum et al. (1999)
k_{aD}	10^6 M ⁻¹ s ⁻¹	Plum et al. (1999)
k_{dH_1}	5.5531×10^{-7} s ⁻¹	Plum et al. (1999)
k_{dH_2}	8.1493×10^{-5} s ⁻¹	Plum et al. (1999)
k_{dU}	≈ 0	Plum et al. (1999)
k_{dD}	10^4 M ⁻¹ s ⁻¹	Plum et al. (1999)
K_m	1.5×10^{-9} M	Sambrook and Russell (2001)
k_{cat}	0.17 M ⁻¹ s ⁻¹	Qiagen (2002)
k_{deg}	1.9×10^{-4} s ⁻¹	Sambrook and Russell (2001)
t_{den}	5 s	PCR protocol
t_{ext}	15 s	PCR protocol

^aUsing nearest-neighbor thermodynamics (Tinoco et al., 2001) and the sequence for 18S (Jayaraman et al., 2000).

RESULTS

Relating DNA Levels and Fluorescence Levels

Most rtPCR analysis methods assume that the efficiency of PCR is constant with respect to cycle number over the range of cycle numbers used for quantitation. Under this assumption, the ratio $\frac{F_n}{F_{n-1}}$ should be constant for a reasonable number of cycles n . An analysis of the data, using Equation (6), from a series of rtPCR measurements using 18S cDNA as the template, shows that $\frac{F_n}{F_{n-1}}$ is *not* constant for each cycle n (Fig. 2). For each starting concentration in this dilution series, the log(Fluorescence) versus cycle number data (Fig. 2a) appears visually to be linear over a range of about eight cycles. In calculating the efficiencies, we see rather that at early cycle numbers, the efficiency is close to 100% but noisy because of the relatively low signals. Then, after only a few more (~ 3) cycles, the efficiency decreases rapidly until it reaches zero. This behavior is not unique to the experimental system used (rat 18S cDNA), as it has been observed by others (Peirson et al., 2003) and for other genes studied in our own laboratory (results not shown). Decreasing the initial template concentration causes a shift in both the fluorescence and the efficiency plots towards higher cycle numbers, with both sets of curves varying in roughly parallel fashion over most cycle numbers. It should also be noted that the shift in the starting cycle of each plot is a result of differentiating meaningful fluorescence levels (actual target DNA) from background noise. For smaller starting template concentrations, the fluorescence levels remain below the sensitivity limit for a larger number of cycles. As a result, the plots for

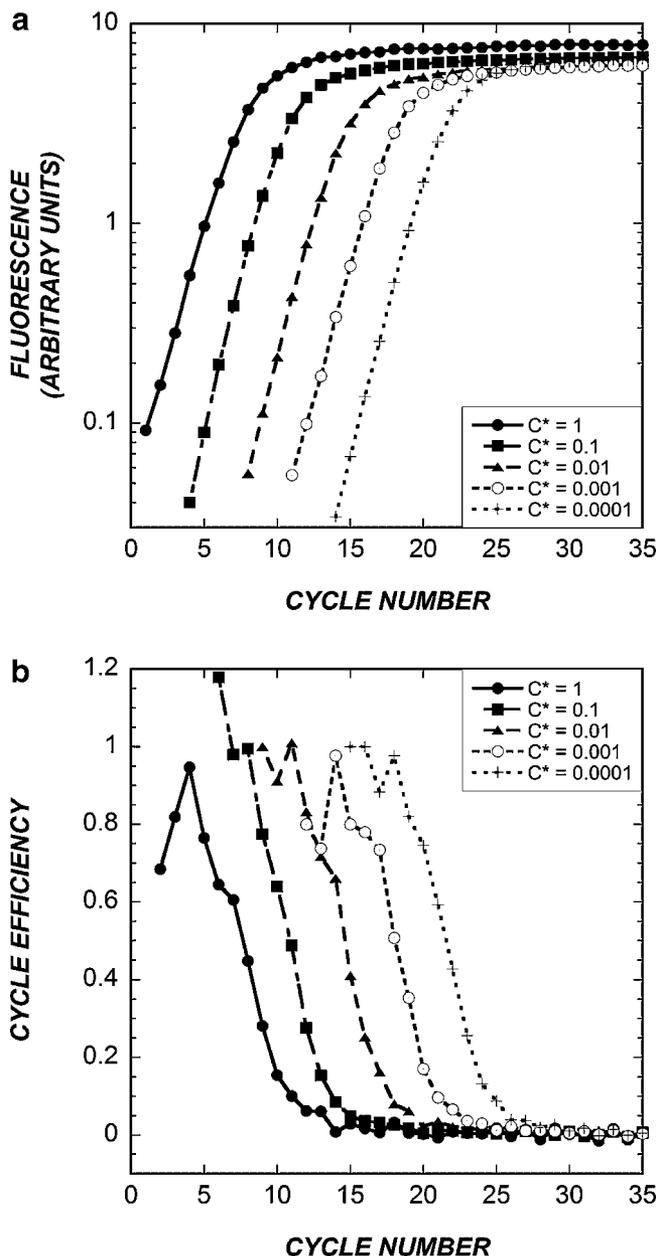


Figure 2. Experimental plots of rtPCR efficiency versus cycle number: (a) background corrected fluorescence values, (b) cycle-dependent efficiencies calculated directly from fluorescence data. Experimental data was generated for a dilution series of rat 18S cDNA relative to a stock concentration, which is indicated as $C^* = 1$.

lower initial template concentrations begin at higher cycle numbers. The difficulty in differentiating meaningful fluorescence levels from background noise is also responsible for the greater variability in efficiency at early cycles.

Efficiency Model

Equilibrium Model of Annealing With Kinetic Model of Extension

The PCR model with annealing reaching equilibrium was solved for a range of initial template concentrations (Fig. 3).

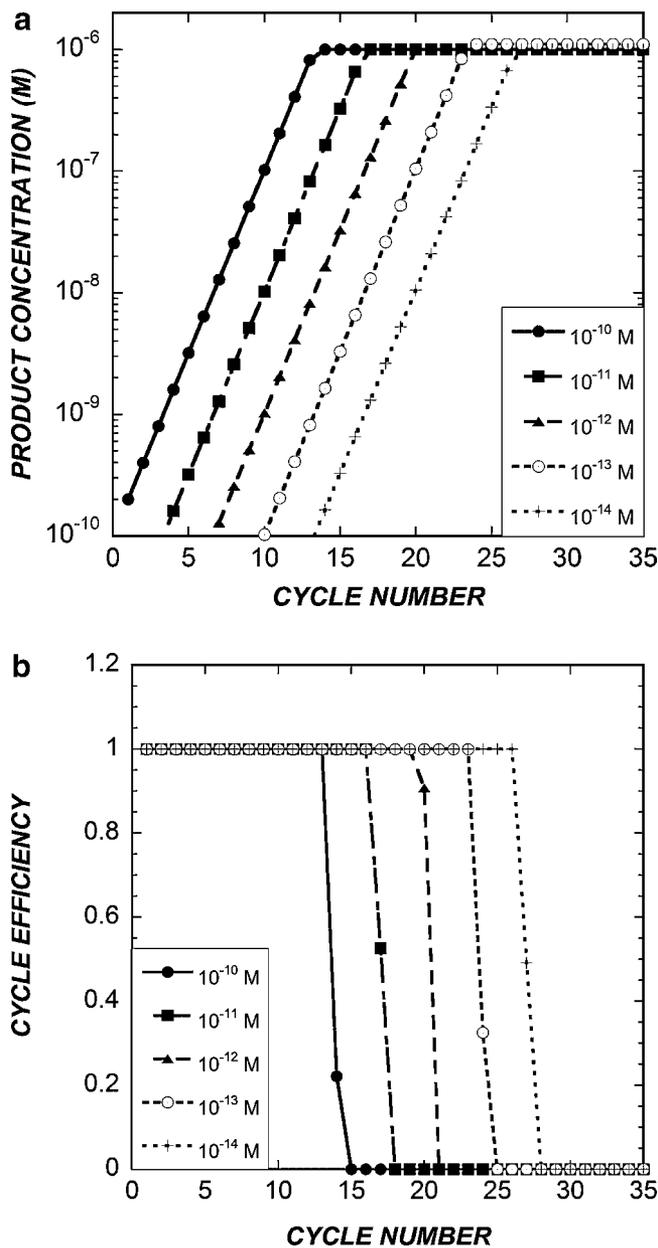


Figure 3. Simulation results of rtPCR over 35 cycles using the equilibrium annealing model and the parameters in Table I: (a) product concentration values, (b) cycle-dependent efficiencies from rtPCR simulations at varying concentration values.

In each case, the efficiency of PCR begins at 100%, and decreases dramatically after a certain number of cycles. The cycle at which efficiency drops off varied with starting template concentration, in the manner observed experimentally (cf. Fig. 2). Although the basic behavior of these simulated plots is similar to that of the experimental plots (efficiency begins at 100% and eventually decreases to 0), the rate of efficiency decrease predicted by the equilibrium model is significantly more abrupt than observed experimentally. In the equilibrium model, template–template reannealing could not be included, and this omission may belie the abruptness of the efficiency transition in this case.

Kinetic Model of Annealing and Extension

Because of the relatively short cycle time, it is possible that the reactions in the annealing step of rtPCR are kinetically limited. Furthermore, template–template reannealing is likely an important factor affecting PCR efficiency. To explore these factors, we developed a mass action kinetic model of the same annealing process (Eqs. 18–25). Using the PCR model with kinetic annealing, the efficiency of PCR remains at 100% for several cycles, and then decreases sharply. However, the rate of decrease is more gradual than that predicted for the thermodynamic annealing step (Fig. 4). Indeed, the pattern of efficiency versus cycle number for the

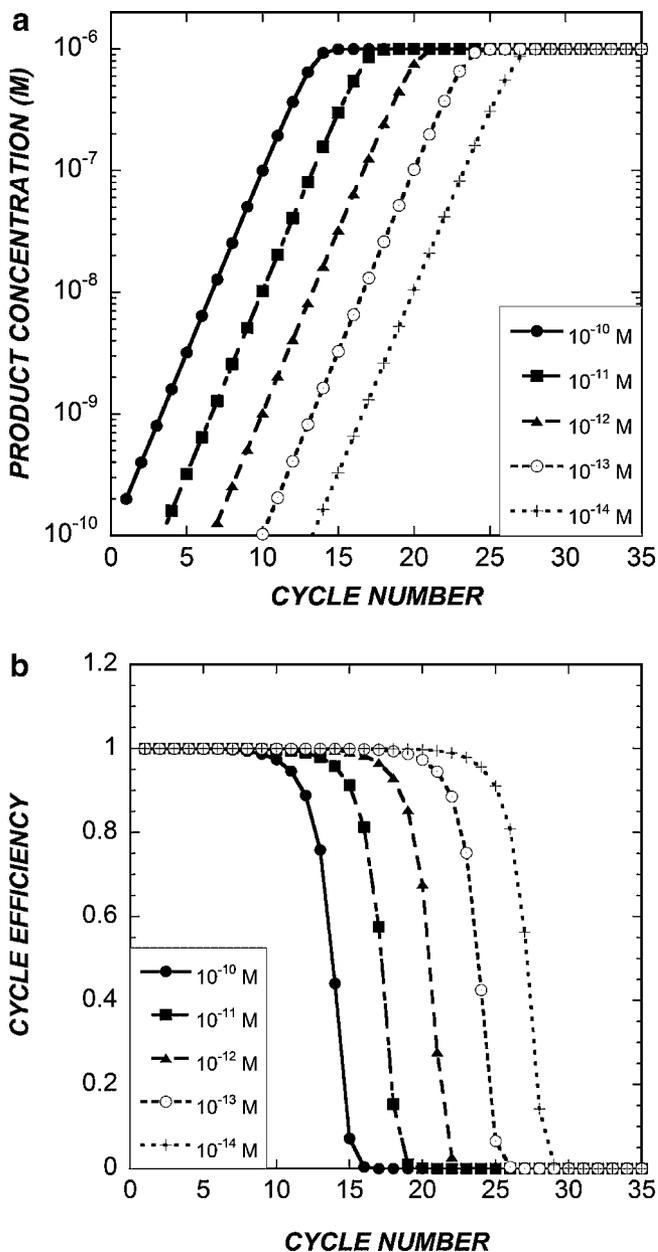


Figure 4. Simulation results of rtPCR over 35 cycles using the kinetic annealing model and the parameters in Table I: (a) product concentration values, (b) cycle-dependent efficiencies from rtPCR simulations at varying concentration values.

kinetic model is closer to that observed experimentally than for the thermodynamics model (cf. Fig. 2). Again, as initial template concentration decreases, the pattern of efficiency versus cycle does not change; rather, each plot is further shifted to the right.

Given the very minor decrease in enzymatic activity with cycle number predicted by the enzyme specifications (Eq. 27), we expect that the decrease in PCR efficiency is due either to the depletion of materials (i.e., primers) in the annealing step or to competition of the primer-template annealing reaction with other reactions, most notably template–template reannealing. Plotting the concentrations of the primer and template–template hybrids concurrently with the efficiency provides some insight into this question (Fig. 5). We find that the efficiency is high while the concentration of primer is much greater than that of the template. Template–template reannealing increases as the square of the template concentration, and, as template concentration approaches that of primer, not enough primer remains to anneal with it. As a result, the efficiency drops, first gradually when the template is within 1–2 orders of magnitude of the primer concentration, then dramatically as the template and primer concentrations are of the same order of magnitude. Product reannealing competitively inhibits primer-template annealing, but this effect lags the primer depletion effect and does not appear to be the primary cause of efficiency loss.

Because the model has a large number of parameters, it is important to consider the effect of their variation on the model predictions. In fact, within the physically realistic range, the particular values had almost no effect on the model predictions. A few examples of parameter variations are shown in Figure 6. For each of the changes tested (hybrid

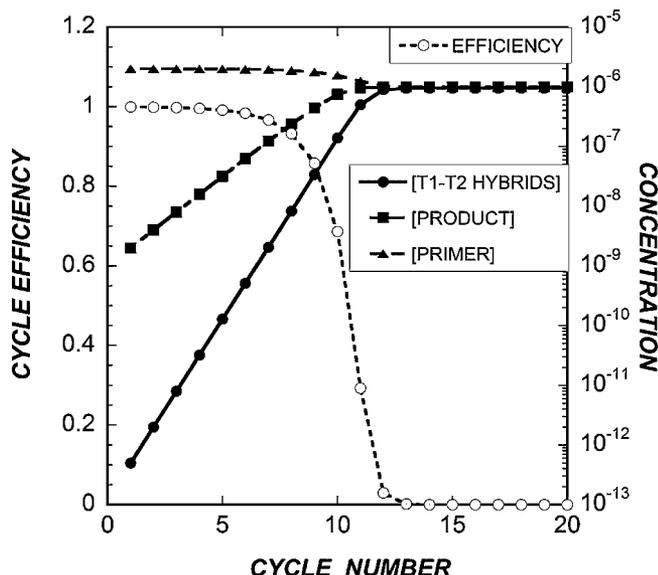


Figure 5. Relationship of efficiency changes to accumulation of product and to disappearance of primers. For the base case parameter values, the efficiency begins to decrease significantly when the product concentration reaches a level on the order of that of the primer concentration. Concentrations of T1–T2 hybrids formed during the annealing step are also shown.

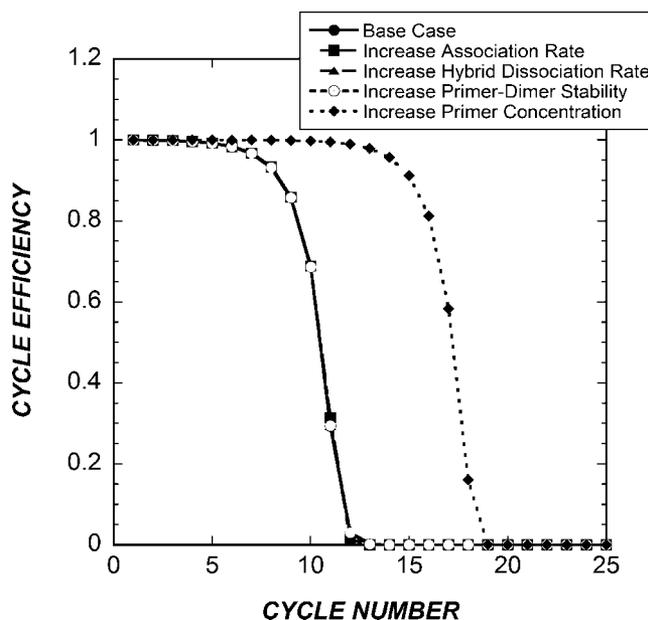


Figure 6. Parametric variation. Several of the model parameters were varied in order to elucidate their effect on model predictions. An increase in association rate reflects changes in each of the hybridization rates $\{k_{aH1}, k_{aH2}, k_{aU}, k_{aD}\}$ from 10^6 – 10^7 /Ms. An increase in hybrid dissociation rate reflects a change in k_{d1} from 5.55×10^{-7} – 5.55×10^{-6} /Ms. An increase in primer-dimer stability reflects a change in k_{dD} from 10^4 – 10^2 /M. An increase in primer concentration was made from 10^{-6} – 10^{-4} M. All the curves except the increase in primer concentration are indistinguishable.

association rates, primer-template dissociation rate, and primer-dimer stability), the new efficiency versus cycle number behavior is indistinguishable from the base case. Thus, it appears that the efficiency is governed primarily by the experimental variables of template concentration and primer concentration, which can be controlled by the practitioner, with a weak dependence on the exact value of the kinetic constants within the experimentally accessible range. Increasing the primer concentration 100-fold resulted in a shift of 6–7 cycles in the onset of efficiency decrease, consistent with the notion that primer concentration, independently or relative to template concentration, is a significant factor controlling the PCR efficiency (Fig. 6).

DISCUSSION

Real-time PCR has gained increasingly widespread use in the past few years as a technique to quantify the levels of specific transcripts in a pool of RNA extracted from cells or tissues. The exponential amplification inherent in the PCR reaction confers sensitivity to the assay, and the “real-time” data acquisition lends itself to automated data analysis, from which quantitative values are obtained. For these reasons, rtPCR has become the method of choice for validating the most significant findings in DNA microarray experiments, and it is used alone for more focused investigations of gene expression in a variety of biological systems. Given the emergence of rtPCR as a major biological analysis technique, a careful examination of the factors influencing its efficiency is important for understanding its power and its limitations.

The quantitation of absolute or relative mRNA levels from rtPCR data is a subject that has received considerable attention in recent years. The variation in threshold cycle number with concentration in dilution series can be used to fit an efficiency that can be considered an average value across all the runs in the series and all cycles within that run. As run-to-run variation in efficiency has been demonstrated to occur (Ramakers et al., 2003), an alternative approach is to obtain a run-specific efficiency from a select group of data points where the amplification appears to be exponential, using Equation (5). The message of Figure 2, however, is that it can be difficult to identify a range of cycle numbers where there is sufficient signal to noise, yet constant efficiency, within an individual PCR run. The determination of such a range using statistical methods can improve this approach (Tichopad et al., 2003), but it is clearly desirable to either optimize the PCR reaction conditions to extend the range of usable cycles or incorporate variation of cycle efficiency into the data analysis. This decrease in efficiency does not appear to be the result of instrument settings or an inner filter effect related to fluorophore concentration, given the very high dilution of SYBR[®] dyes used for quantitation and the small path length (~ 1 cm) of the rt-PCR capillaries. Absent an unequivocal explanation, we developed a mathematical model to aid in our interpretation of these results.

A decrease in cycle efficiency, particularly at high-cycle numbers, has been discussed and modeled previously in the context of conventional PCR. Previous models focused mainly on the decrease in enzyme activity with repeated exposure to high temperatures in the denaturation step (Hsu et al., 1997) and on the saturation of enzyme activity with high levels of substrate (primer-template hybrids) (Schnell and Mendoza, 1997). Additional factors implicated in the saturation of PCR product formation include primer-dimer formation, the competitive binding of DNA polymerase to its amplification products, and depletion of NTPs (Halford et al., 1999; Kainz, 2000). Among these, the potential for primer-dimer formation is certainly an important consideration in primer design, but it is not likely to be a factor in the cycle-dependent efficiency behavior, since the amount of primer actually decreases with increasing cycle number. Nonetheless, primer-dimer formation is included in our model. We also incorporated explicitly the cycle dependence of enzyme activity into our simulations and found that it contributed negligibly to efficiency decline. This is not inconsistent with the strong role for enzyme deactivation implied in previous studies (Hsu et al., 1997), as enzymes of greater thermostability are available currently, and the rapid cycling in a rtPCR protocol further minimizes this effect as compared to conventional PCR. The depletion of dNTPs could, in principle, affect PCR either if they are stoichiometrically limiting, or if they are low enough to be incorporated inefficiently by the Taq polymerase enzyme. However, neither of these mechanisms seems to be a significant factor, since the starting concentration (e.g., 400 μM for each dNTP (Qiagen, 2002)) exceeds that consumed over the entire course of a PCR run, which starts with 1 μM of

primers. Furthermore, the Michaelis constant of the Taq polymerase for dNTPs is ~ 10 μM , which is much less than the concentration of dNTPs.

Since other mechanisms cited do not seem to play a major role in explaining the efficiency behavior that is commonly observed in rtPCR, we focused our attention on the competitive binding events occurring during the annealing phase. At the temperatures typically employed for annealing (55–60°C), both primer-dimer formation and template–template reannealing may compete with primer-template annealing. Thermodynamically, template–template reannealing is always favored, since the greater length of duplex formation between complementary templates results in much higher affinity than for primer-template hybrid formation. However, at the early stages of PCR, the primer concentration is many orders of magnitude greater than that of the template, and so template–template reannealing is minimal, and 100% efficiency can be achieved. We found that as the template concentration rose to within 1–2 orders of magnitude of the primer concentration, primer-template annealing was no longer so strongly favored by mass action and decreased substantially. The use of an equilibrium annealing step was based on the annealing time of an rtPCR cycle (e.g., 20 s) being considerably greater than the relaxation time, τ , for bimolecular association of primer and template,

$$\tau = \frac{1}{k_{dH_1} + k_{aH_1}([P_1]_{eq} + [T_1]_{eq})} \quad (39)$$

where the subscript eq denotes equilibrium values. The value of the relaxation time is ~ 1 s. Using this approach, model simulations correctly predict the effect of starting template concentration on fluorescence intensity curves and the decline in PCR efficiency induced at higher cycle numbers (Fig. 3). However, the decline in efficiency predicted using the equilibrium-annealing model was in fact more precipitous than observed in experimental data.

As a result, we found that we needed to incorporate the competitive annealing events in a kinetic reaction framework. A key element of the kinetic model is the ability to include template–template reannealing as an irreversible reaction. Using the kinetic model, we found an earlier and more gradual decline in efficiency with cycle number, more closely resembling the behavior observed experimentally. Even with the molecular events in the annealing stage handled kinetically, our model somewhat over-predicts the rate of efficiency decline. Several factors not accounted for explicitly in the model may be responsible for this over-prediction. These include stabilization of the primer-template duplex by polymerase binding, the possibility of forming nonspecific primer-enzyme complexes, minor depletion effects due to the dNTP concentration, or interference from the DNA-binding dyes used for real-time detection (Roth, 2002).

Probe design is a critical factor in the success of an rtPCR experiment. Different probes would be reflected in the model in various ways. The values of the dissociation rate constants were estimated using equilibrium constants computed via

nearest-neighbor thermodynamics (Tinoco et al., 2001) and an assumed association rate constant of $10^6 \text{ M}^{-1} \text{ s}^{-1}$, which is typical for nucleic-acid hybridization (Plum et al., 1999). The sequences of different probes would change the values of these parameters somewhat, but the model predictions were quite robust to large change in the values of these parameters. The selection of probes dictates the amplicon (template) size, and this would affect the effective k_{cat} used in the extension phase of the reaction. Using our assumed length of 200 bp, the extension time was sufficient for complete conversion. However, longer amplicons could result in a limitation in extension and increase the potential for dNTP depletion. A feature of probe design that is difficult to incorporate into the model is probe specificity. If the PCR probe is binding to other species in the reaction mixture, less will be available for binding to template, and a decrease in efficiency would result. In a similar vein, it should be noted that our model envisions detection using DNA-binding dyes, such as the SYBR[®] dyes. If a hybridization probe with dyes attached is used, this presents another nucleic acid that may interact with the primer, the template or the polymerase enzyme. For a full treatment of this detection format, these interactions would need to be incorporated into the model, but doing so is beyond the present scope.

The ability to quantify the efficiency of a specific rtPCR experiment allows for the development of methods that can more accurately determine gene expression levels from fluorescence intensities. At a fundamental level, our model provides a framework to quantify the multifactorial effects of a change in PCR protocol on the quantitation of a target mRNA. Importantly, the lack of sensitivity of model outputs to variation of model parameters suggests that once a suitable primer-template sequence has been identified, a minimum of effort need be expended in tweaking reaction conditions, as such changes negligibly alter the PCR reaction kinetics (Fig. 6). However, there are several other experimental variables that the model suggests could be used to extend the high efficiency range. For example, increasing the primer concentration is predicted to result in high efficiency product formation for a greater number of cycles without extending the range of sub-detectable product concentrations. This would need to be balanced against detrimental effects of increasing primer concentration, which include primer-dimer formation, non-specific product formation, and primer-enzyme interactions. Since the annealing time is roughly 20-fold greater than the relaxation time for primer-template annealing, decreasing the annealing time may also be a feasible approach to improve efficiency.

Alternatively, calculation of efficiencies could be used in determination of the threshold fluorescence value in conventional quantitation techniques. In principle, we can calculate the efficiencies at each cycle and use all of these values to back-calculate the starting template concentration, using Equation (7). As this is somewhat cumbersome, we are investigating alternatives that parameterize the cycle-dependent behavior in a way that is consistent with our observations. In the near future, we expect that the mechanistic molecular

insights gained from this study will translate into more accurate quantitation of rtPCR data.

J.L.G. acknowledges the Rutgers Mathematics Department REU program for supporting her participation in this research. Further support was provided by an NSF CAREER Award (BES-0238617) to C.M.R. The authors thank Anthony Nicolini for help with parameter estimation and an anonymous referee for unusually detailed and insightful comments. C.M.R. thanks Arul Jayaraman for many fruitful discussions over the years about rtPCR.

References

- Freeman WM, Walker SJ, Vrana KE. 1999. Quantitative RT-PCR: Pitfalls and potential. *Biotechniques* 26(1):112–122, 124–125.
- Goidin D, Mamessier A, Staquet MJ, Schmitt D, Berthier-Vergnes O. 2001. Ribosomal 18S RNA prevails over glyceraldehyde-3-phosphate dehydrogenase and beta-actin genes as internal standard for quantitative comparison of mRNA levels in invasive and noninvasive human melanoma cell subpopulations. *Anal Biochem* 295(1):17–21.
- Halford WP, Falco VC, Gebhardt BM, Carr DJ. 1999. The inherent quantitative capacity of the reverse transcription-polymerase chain reaction. *Anal Biochem* 266(2):181–191.
- Hsu JT, Das S, Mohapatra S. 1997. Polymerase chain reaction engineering. *Biotechnol Bioeng* 55:359–366.
- Jayaraman A, Yarmush ML, Roth CM. 2000. Dynamics of gene expression in rat hepatocytes under stress. *Metabolic Eng* 2:239–251.
- Kainz P. 2000. The PCR plateau phase—Towards an understanding of its limitations. *Biochim Biophys Acta* 1494(1-2):23–27.
- Liu W, Saint DA. 2002. A new quantitative method of real time reverse transcription polymerase chain reaction assay based on simulation of polymerase chain reaction kinetics. *Anal Biochem* 302(1):52–59.
- Livak KJ, Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C(T)}$ method. *Methods* 25(4):402–408.
- Peirson SN, Butler JN, Foster RG. 2003. Experimental validation of novel and conventional approaches to quantitative real-time PCR data analysis. *Nucleic Acids Res* 31(14):e73.
- Pfaffl MW. 2001. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res* 29(9):e45.
- Plum GE, Breslauer KJ, Roberts RW. 1999. Thermodynamics and kinetics of nucleic acid association/dissociation and folding processes. *Comprehensive natural products chemistry*. Oxford, UK: Elsevier Science Ltd. 7, ch. 2:15–33.
- Qiagen. 2002. HotStar Taq PCR Handbook. Valencia, CA.
- Ramakers C, Ruijter JM, Deprez RH, Moorman AF. 2003. Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. *Neurosci Lett* 339(1):62–66.
- Roth CM. 2002. Quantifying gene expression. *Curr Issues Mol Biol* 4(3):93–100.
- Rutledge RG, Cote C. 2003. Mathematics of quantitative kinetic PCR and the application of standard curves. *Nucleic Acids Res* 31(16):e93.
- Sambrook J, Russell DW. 2001. *Molecular cloning: A laboratory manual*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.
- Schmittgen TD, Zakrajsek BA. 2000. Effect of experimental treatment on housekeeping gene expression: Validation by real-time, quantitative RT-PCR. *J Biochem Biophys Methods* 46(1-2):69–81.
- Schnell S, Mendoza C. 1997. Theoretical description of the polymerase chain reaction. *J Theor Biol* 188(3):313–318.
- Stagliano KE, Carchman E, Deb S. 2003. Real-time polymerase chain reaction quantitation of relative expression of genes modulated by p53 using SYBR green I. *Methods Mol Biol* 234:73–91.
- Tichopad A, Dilger M, Schwarz G, Pfaffl MW. 2003. Standardized determination of real-time PCR efficiency from a single reaction set-up. *Nucleic Acids Res* 31(20):e122.
- Tinoco I, Sauer K, Wang JC, Puglisi JD. 2001. *Physical chemistry: Principles and applications in biological sciences*. Upper Saddle River, NJ: Prentice Hall.