PCR Model 2

Ben Shirt-Ediss

Newcastle University UK: August/September 2017

PCR Cycle - Recap

In the very hot **denaturation phase**, all DNA complexes become single-stranded. On entering the relatively cool **annealing phase** afterwards, some amplicon single strands will hybridise with primers instead of complementary amplicons. Those hybridising with primers can attract a DNA polymerase, which starts slowly turning the ternary complex into dsDNA. As the complexes convert to dsDNA, they gradually become more thermally stable. Then, when the temperature is increased in the final **extension phase**, the partly elongated complexes survive this temperature jump, and elongate even more quickly, as the hot temperature is optimal for the DNA polymerase enzyme processivity). If the amplicon is several Kb, the elongation phase is necessary to quickly template-copy all of it (just using the annealing phase would be too slow).

Problems with Model 1

Model 1 had full primer extension in 1 step (!). This caused a problem: ternary complexes in the **annealing phase** either melted back to ssDNA, or were instantly transformed into super-stable dsDNA amplicons. Then, when the **extension phase** came, the remaining ternary complexes – which had not melted or had not been extended at all – were even more likely to melt at the hotter temperature. The tiny fraction of ternary complexes surviving into extension were therefore insignificant, and the extension phase was redundant: the type of enzyme used, its speed and the length of the extension phase had no effect on the results.

Improvements in Model 2

Two approaches can be taken to make the extension phase more functional:

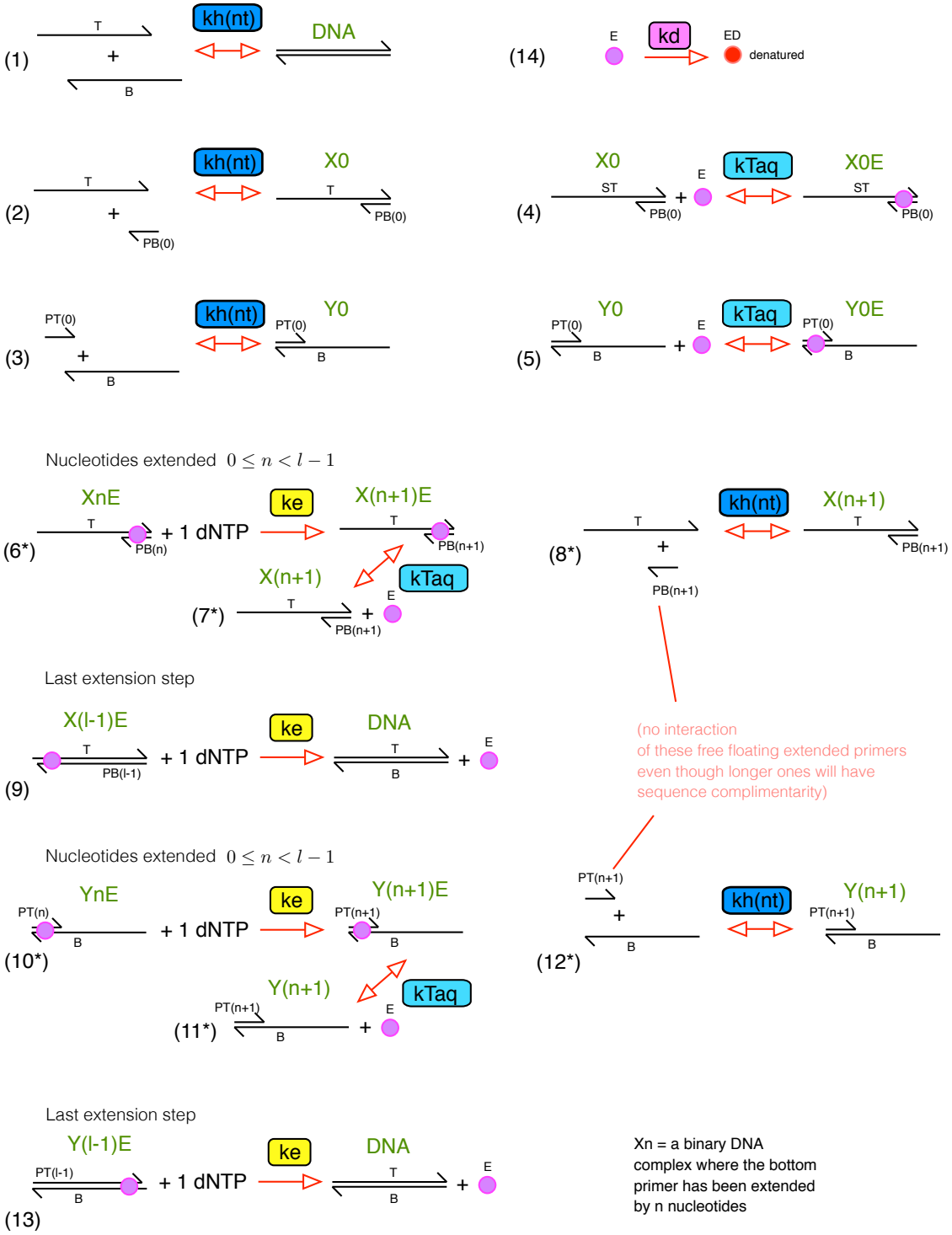
- **Approach 1:** consider all individual steps of primer extension, and consider that the complexes can melt into a full strand and an intermediately extended strand. BUT, don't count interactions between the intermediate-extended strands themselves. Counting these interactions causes a combinatorial explosion in the number of possible reactions in the system. Furthermore, it is hard to discern if some of these reactions would be viable or not (i.e. what can a polymerase stick to? does ssDNA secondary structure kinetically prohibit some hybridisation bindings? etc..). Instead, I consider that only fully extended strands can hybridise, and interact with primers.
- **Approach 2:** coarse-grain the steps of primer extension, reducing 80 to around 10 (or less) extension steps, and explicitly count interactions between all the intermediate-extended strands formed (these chimera complexes can, themselves, be extended). This coarse-graining requires (i) deriving rate expressions for transitions between the coarse-grain steps (overall rate law for consecutive sequence of first-order irreversible reactions) and (ii) automatically deriving all reactions e.g. via linear sequence matching. Step (ii) is not so trivial.

I will take Approach 1 for now, as it is simplest for the DoE 2017 workshop. Approach 2 could be published later as a first approach to partial extension modelling in the PCR reaction.

Model 2 also has improved thermodynamics for the polymerase association/disassociation reaction.

1 Reaction Diagram

PCR Model 2 -- hidef extension added
(but no interaction of partially extended products)



2 Reactions (for 1 nt extension)

$$T + B \rightleftharpoons^{kh} DNA \tag{1}$$

$$T + PB0 \rightleftharpoons^{kh} X0 \tag{2}$$

$$PT0 + B \rightleftharpoons^{kh} Y0 \tag{3}$$

$$X0 + E \rightleftharpoons^{kTaq} X0E \tag{4}$$

$$Y0 + E \rightleftharpoons^{kTaq} Y0E \tag{5}$$

Extension steps for X complex where nucleotides extended $0 \leq n < l - 1$

$$XnE + dNTP \rightarrow^{ke} X(n+1)E \tag{6}$$

$$X(n+1) + E \rightleftharpoons^{kTaq} X(n+1)E \tag{7}$$

$$T + PB(n+1) \rightleftharpoons^{kh} X(n+1) \tag{8}$$

Final extension step for X complex

$$X(l-1)E + dNTP \rightarrow^{ke} DNA + E \tag{9}$$

Extension steps for Y complex where nucleotides extended $0 \leq n < l - 1$

$$YnE + dNTP \rightarrow^{ke} Y(n+1)E \tag{10}$$

$$Y(n+1) + E \rightleftharpoons^{kTaq} Y(n+1)E \tag{11}$$

$$PT(n+1) + B \rightleftharpoons^{kh} Y(n+1) \tag{12}$$

Final extension step for Y complex

$$Y(l-1)E + dNTP \rightarrow^{ke} DNA + E \tag{13}$$

Enzyme decay

$$E \rightarrow^{kd} ED \tag{14}$$

Later: In the model, I found 1 nt per time extension to execute too slowly. So, I implemented extension where multiple nt were extended per time. I treated each extension jump as following MAK kinetics, although this would not be the case in reality, as adding multiple dNTPs would involve many sequential reactions. However, it worked satisfactorily. See code in g_pcrsim.py.

3 Rate Constants

3.1 k_h reactions

These are reversible DNA hybridisation/melting reactions. The kinetic forward rate is fixed, regardless of hybridisation length:

$$k_f = 10^6 M^{-1} s^{-1}$$

and the reverse rate is variable, calculated so that the reaction has the correct equilibrium constant:

$$K_{eq} = \frac{k_f}{k_r}$$

$$k_r = \frac{k_f}{K_{eq}} \qquad s^{-1}$$

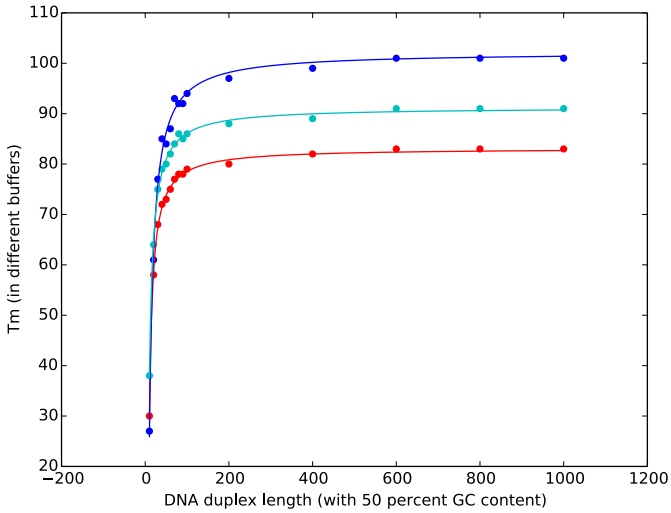
Van't Hoff's formula is used to calculate the equilibrium constant from thermodynamic parameters:

$$K_{eq} = e^{\left[-\frac{\triangle G^0}{RT}\right]} = e^{\left[-\frac{\triangle H^0 - T\triangle S^0}{RT}\right]} = e^{\left[-\frac{\triangle H^0}{RT} + \frac{\triangle S^0}{R}\right]}$$

Units
In this model, $\triangle H^0$ and $\triangle S^0$ are in $cal\ mol^{-1}$, Gas constant R is in $cal\ mol^{-1} K^{-1}$ and temperature is in K.

Thermodynamic parameter $\triangle S^0 = 300\ cal\ mol^{-1}$ is presumed fixed, and $\triangle H^0$ depends on the hybridisation length.

To calculate $\triangle H^0$, firstly the melting temperature for the DNA duplex is obtained from a best-fit to the NEB melting temperature calculator. The NEB melting temperature calculator allows calculation of T_m in both Taq (red, below) and Phusion (blue, below) buffers:



This T_m value is then used to calculate $\triangle H^0$. From melt-curve math:

$$T_m = \frac{\Delta H_0}{\Delta S_0 - Rln\left[\frac{4}{C_T}\right]}$$

Therefore:

$$\Delta H_0 = T_m\left(\Delta S_0 - Rln\left[\frac{4}{C_T}\right]\right)$$

where NEB use a total single strand concentration $C_T = 500nM$.

3.2 k_{Taq} reactions

These are reversible reactions whereby the DNA polymerase associates/disassociates from a primed complex (X or Y).

According to 2003 paper "Thermodynamics of the binding of Thermus aquaticus DNA polymerase to primed-template DNA" (DOI: 10.1093/nar/gkg774), the ΔG^0 for Taq polymerse binding to primed DNA (a 63/70 mer complex) is constant at around $-11\ kcal\ mol^{-1}$ over 5-70 degrees celsius.

2015 paper "Polymerase/DNA interactions and enzymatic activity: multi- parameter analysis with electro- switchable biosurfaces" (DOI: 10.1038/srep12066) states that ΔG^0 for Taq polymerse binding to primed DNA (a 36/54 primer/template complex) is $-59\ kJ\ mol^{-1} = -14.1\ kcal\ mol^{-1}$ at 25C, agreeing well with the estimate above. They also measure the kinetic on-rate of Taq to 36/54 primed DNA complexes as $1.9 \times 10^7\ M^{-1}s^{-1}$ at 25C.

From these data, I assume:

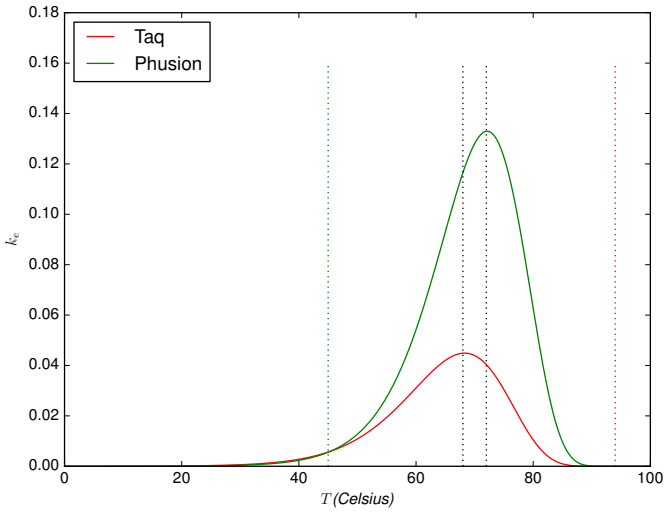
- Taq and Phusion polymerases have identical thermodynamic and kinetic on-rate behaviour
- $\Delta G^0 = -11\ kcal\ mol^{-1}$ for polymerase association with a primed complex, over the entire temperature range (5-100C)
- $k_f = 1.9 \times 10^7\ M^{-1}s^{-1}$ is the polymerase on-rate, over the entire temperature range (5-100C)
- k_r, the polymerase off rate, is calculated from obeying the equilibrium constant $k_r = \frac{k_f}{K_{eq}}$ where $K_{eq} = e^{\left[-\frac{\triangle G^0}{RT}\right]}$

3.3 k_e reactions

These are irreversible reactions whereby a primed complex extends by 1 more nucleotide. In this model, it is an educated guess how rate k_e depends on temperature for the different polymerases. A log-normal distribution reflected in the y-axis is arbitrarily chosen to model the dependence of the rate constant k_e with temperature:

$$k_e = HV\frac{1}{S\sqrt{2\pi}(100 - T_C)}e^{-\frac{(ln(100-T_C)-M)^2}{2S^2}}$$

Where T_C is the temperature in Celsius.



The V, M and S constants control the curve shape and position, and are chosen differently for the Taq and Phusion polymerases such that:

- Taq has an optimal elongation temperature of 68C
- Phusion has an optimal elongation temperature of 72C
- Phusion elongates 3 times faster than Taq, when the enzymes are at their optimal temperature. This is based on the rough statement that "Taq elongates at 1kb per min, and Phusion elongates at 1kb every 20 seconds" (about 3kb per minute). The constant V maintains the 3x ratio between curve heights.

	Taq	Phusion
H	same	same
V	1	2.6
M	3.53	3.4
S	0.27	0.27

Parameter H scales the height of both curves in the same ratio. This parameter can be tuned to give a k_e such that when Taq polymerase is used, and the amplicons are 1Kb long, then waiting 1 minute in the extension phase is sufficient to elongate the majority of the amplicons fully.

3.3.1 Amendment to extension reaction rate

In model 1, the reaction rate of the 1-nt extension reaction was

$$k_e[SnE][dNTP]$$

where S may be an X or Y complex, and $[dNTP]$ is the total concentration of the dNTP pool.

However, this is not strictly correct, since there are 4 types of dNTP, and the polymerase will only incorporate the type which is complementary to the current target nucleotide – not all types of dNTP.

A better expression for the rate, assuming that the polymerase depletes all 4 types of dNTP at approximately equal rate is:

$$k_e[SnE]\frac{[dNTP]}{4}$$

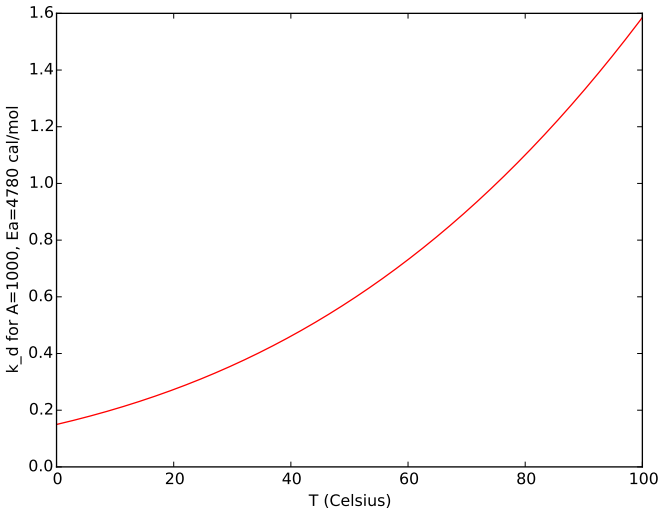
Here, only the concentration of the currently complementary dNTP affects the reaction rate.

3.4 k_d reaction

This reaction implements the irreversible decay of the polymerase, which increases with increasing temperature. Decay rate constant k_d is thus assumed to have Arrhenius-like behaviour. It is calculated by

$$k_d = Ae^{-\frac{E_a}{RT}} \qquad s^{-1}$$

Where E_a is the activation energy barrier from a functioning polymerase state, to a denatured polymerase state. Both the Taq and Phusion polymerases are assumed to have the same denaturation curve.



The E_a constant is fixed, and the A multiplier is tuned – like above – such that the follwing overall macroscopic observation is reproduced: "The normal half life for the enzyme is of the region of >40 cycles of 1min 96C, 1min 55C, 1min 72C" (http://www.bio.net/bionet/mm/methods/1995-June/029361.html). Thus, A is set such that after 40 cycles of the above, we end up with half the amount of enzyme we started with.