6-1-2006

Kinetics of the DNA polymerase *pyrococcus kodakaraensis*

Mark A. Griep
University of Nebraska-Lincoln, mgriep1@unl.edu

Casey A. Kotera
University of Nebraska - Lincoln

R.M. Nelson
University of Nebraska - Lincoln

Hendrik J. Viljoen
University of Nebraska - Lincoln, hviljoen1@unl.edu

Kinetics of the DNA polymerase *pyrococcus kodakaraensis*

M. A. Griep [a], C. A. Kotera [b], R. M. Nelson [b], and H. J. Viljoen [c],*

[a] *Department of Chemistry, University of Nebraska–Lincoln, Lincoln, NE 68588*
[b] *Department of Chemical Engineering, University of Nebraska–Lincoln, Lincoln, NE 68588*
[c] *Megabase Research Products, Lincoln, NE 68504*

* Corresponding author. E-mail: hviljoen1@unl.edu

Abstract: The polymerase chain reaction is one of the most important reactions in molecular biology. Single stranded DNA is copied in a complex series of steps, at the core of which lies the action of the DNA polymerase. At each nucleotide along the template, the polymerase screens the dNTP pool until it finds the complementary dNTP. The insertion of each dNMP is a balance between high fidelity and rapid elongation. In this study the kinetics of the β type polymerase *pyrococcus kodakaraensis* (KOD) is analyzed. The kinetics is influenced by reaction conditions such as the dNTP pool composition and temperature. In a previous study by Viljoen et al. [2005, A macroscopic kinetic model for DNA polymerase elongation and high-fidelity nucleotide selection. Computational Biology and Chemistry 29, 101–110], a macroscopic kinetics expression of the polymerase chain reaction has been derived. The model contains four parameters that are intrinsic to a specific polymerase. The experiments to measure the temperature-dependence of the parameters for KOD DNA polymerase are reported. The results indicate that the optimal temperature for an equimolar dNTP pool is 72.5 °C and the optimum temperature shifts to lower temperatures when the dNTP pool composition is biased.

Keywords: Kinetics, Polymerase chain reaction, Mathematical model, Temperature, dNTP pool

1. Introduction

The polymerase chain reaction (PCR) is a technique to amplify the amount of a specific DNA segment. The denatured segment consists of two single stranded DNA parts (ss-DNA) and they serve as templates to produce to new double stranded DNA segments (dsDNA). Two short DNA fragments (primers), which are complementary to the 3′ ends of the two strands of the segment, anneal to the ssDNA strands and the polymerase catalyzes the extension of the primers. Specific template-directed nucleic acid synthesis is one of the greatest biochemical discoveries (Grundberg-Manago et al., 1955 and Lehman, 2003). The accuracy, relative ability to bypass lesions and ability to elongate polymers processively makes these enzymes crucial to understanding such biological processes as infection, cancer and aging and also identifies them as important drug targets (Ahmed and Tollefsbol, 2003 and Miura and Izuta, 2004). Finally, template-directed nucleic acid polymerases are used in many procedures basic to modern biotechnology including PCR-based recombinant DNA technology, diagnostics and DNA sequencing (Smith, 1980 and Saiki et al., 1985).

PCR is the only technique in molecular biology that uses three different temperatures, called melting, annealing, and elongation, to control the rates of the many reactions that take place. At the melting temperature of 90–95 °C, the DNA or RNA melts to its fully single-stranded state. The temperature is not held there long so that the DNA polymerase does not become irreversibly denatured. At the annealing temperature of 55–70 °C (von Ahsen et al., 2001, Benita et al., 2003 and Owczarzy et al., 2004), the primers form complementary base pairs with their target sequences on the template. This temperature is slightly above the melting temperature for the primer–template complex so that only the most stable base-pairing can occur. This is also the temperature at which the DNA polymerase binds to the primer–template complex and places its active site near the 3′-hydroxyl of the primer (Datta and Li-Cata, 2003). At temperatures below the primer melting temperature (T_m), less canonical base-pairing occurs and, during the elongation phase, leads to the amplification of non-target-
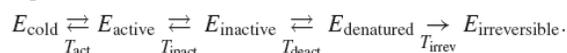
ed sequences and reduced specificity. At temperatures above the T_m, primer base-pairing is more specific but less of it binds causing enhanced specificity with reduced efficiency. At the commonly recommended elongation temperature of 72 °C (Arezi et al., 2003), the DNA polymerase adds nucleotides to the end of the primer in a template-directed manner. It elongates the primer until it stalls or runs out of template.

The PCR step that we are addressing in our macroscopic kinetic model is the elongation step. Like all enzymatic processes, the temperature-dependence of the DNA polymerase exhibits roughly an activity doubling as temperatures rise every 10 °C until it reaches a maximum. After that temperature, called the optimum temperature, the activity falls much more rapidly than it had risen. Most PCR reactions use the optimum temperature as the elongation temperature. This may not be the most desirable PCR temperature, however. Presumably the error rates are lowest when the DNA polymerase has a less flexible structure so that it is less likely to misinsert or misincorporate nucleotides. The trade-off is that the Arrhenius effect will favor higher elongation rates at higher temperatures. To understand the temperature dependence of the enzyme, we need to consider the basis for the observed optimum temperature.

The thermophilic DNA polymerase 1 from *Thermus aquaticus* (Taq DNA pol) is the prototype for understanding the multithermal PCR process. Its use in PCR (Saiki et al., 1988) replaced the mesophilic DNA polymerase 1 from *Escherichia coli*, which had to be added fresh to the reaction after each cycle because it would irreversibly denature during the 95 °C melting step (Saiki et al., 1985). When using Taq DNA pol, the PCR yield roughly triples for every 10 °C increase until it reaches an optimum temperature of 72–75 °C (Innis et al., 1988). In retrospect, this is a curious observation since it has recently been shown that this enzyme unfolds only when the temperature reaches 95 °C (Karantzeni et al., 2003). This enzyme's long half-life at the PCR melting temperature is the consequence of its very high barrier to unfolding. This suggests that there is an inactive enzyme state that is still folded between 75 °C and 95 °C.

When the temperature of a protein solution is raised above its folding temperature, it may irreversibly denature according to a first-order process (Lumry and Eyring, 1954). This would result in the loss of a small percentage of functional enzyme after each PCR cycle. Many proteins become irreversibly denatured due to microaggregates that form due to non-native hydrophoblic interactions or due to mixed disulfides that form between newly exposed cysteines (Klibanov and Ahern, 1987). In addition to these well-studied phenomena, the side chains of folded proteins are constantly subjected to processes such as aspartate backbone hydrolysis, asparagines deamination, and cysteine oxidation (Ahern and Klibanov, 1988), each of which is temperature-dependent. When the protein is folded, these phenomena are fastest for the surface-exposed residues including those at the active site. When the protein is unfolded, the rates for the previously buried residues increase. Therefore, these microchemical changes may denature the enzyme when it is in any one of its folded or unfolded states.

Daniel and co-workers Daniel et al., 2001 and Peterson et al., 2004 recently presented the theory and evidence that thermal optima for monomeric, single-domain enzyme activities result from a rapid (sub-second), temperature-dependent equilibrium between the active and inactive forms of an enzyme, as well as primer-annealing efficiency coupled to enzyme–substrate binding. They equate the inactive form with the thermally unfolded conformation. The case of Taq DNA pol and other palm-type polymerases will be more complex given that they are multi-domain and some have multiple activities. Given that the thermal inactivitation and unfolding of thermophilic polymerases are two separate phenomena, the various equilibria can be represented in the following scheme. As mentioned in the previous paragraph, each state will have its own temperature-dependent, first-order rate of irreverisible denaturation.

$$E_{\text{cold}} \underset{T_{\text{act}}}{\rightleftarrows} E_{\text{active}} \underset{T_{\text{inact}}}{\rightleftarrows} E_{\text{inactive}} \underset{T_{\text{deact}}}{\rightleftarrows} E_{\text{denatured}} \underset{T_{\text{irrev}}}{\rightarrow} E_{\text{irreversible}}.$$

The temperature optimum of about 72 °C must be a balance between the Arrhenius activation energy (E_{cold} to E_{active}) and some thermal inactivation equilibrium (E_{active} to E_{inactive}). For thermophilic polymerases, inactivation is distinct from protein unfolding (E_{inactive} to $E_{\text{denatured}}$). The inactive enzyme could be a conformation in which its "finger" and "thumb" domains are unable to clamp onto the DNA because large thermal motion prevents the various domains from binding one another even though the tertiary structure is intact. In this study, we show how it is possible for polymerases to exhibit a thermal optimum unrelated to thermal unfolding, instead of is a consequence of thermal increases in the time spent in wasteful loops.

The macroscopic kinetic model suggests a different mechanism to account for thermal inactivation of thermophilic polymerases. The polymer extension rate is a balance of productive and unproductive microscopic steps. The temperature-dependence of each step is described by the Arrhenius relationship. After binding one correct dNTP from the pool of four types, the productive steps are rapid and sequential. Upon binding one of the three incorrect dNTPs, there are several routes that might lead to its release before the pool can be sampled again. Thermal inactivation occurs at temperatures where these unproductive steps are faster than the productive ones. In addition, the stability of the enzyme/template complex changes with temperature. The concept of processivity enjoys several definitions, but in essence it is a measure of the number of nucleotides that a polymerase can insert, before the enzyme detaches from the template. The ability of the enzyme to remain bound to the template depends mostly on the temperature. Thus it is expected that the number of nucleotides that may be inserted between detachment events will decrease if the temperature is raised. The enzyme becomes less efficient if it intermittently detaches from the template and needs to attach before processing can continue. Clearly a temperature is reached where gains in the insertion rates are off-set by more frequent detachment/attachment events. A macroscopic kinetic model describes the competition between the different processes and the effect that reaction conditions like temperature and composition may have on the overall extension rate.

The objective of the macrokinetics model is to capture the details of the nucleotide selection, insertion and incorporation in a form that contains the process and control variables, because that is the level at which the user interfaces with the PCR process. The derivation of the model is based on an approach described by Ninio (1987). The behavior of a single enzyme/template complex is tracked over time to determine its average behavior. An ergodic argument is used to show equivalence between the average behavior of a single complex over a long period of time and the average behavior of many complexes at any specific moment in time. The macroscopic model enables us to compare the performance of different polymerases, investigate the roles of temperature and pool compositions and optimize PCR experiments.

2. A brief review of the macroscopic kinetic model

To bridge the gap between the microscopic and macroscopic in the realm of DNA polymerases, we have developed a model for DNA synthesis that accounts for the template nucleotide sequence and dNTP pool to predict DNA synthesis rates and error frequencies (Whitney et al., 2004 and Viljoen et al., 2005). Whereas a microscopic model of PCR focuses on a complex sequence of steps, a macroscopic model collects all detail into a few key steps. Although this reduction in complexity is accompanied a reduction in detailed knowledge, the rate-limiting step can still be identified and further data become largely superfluous. The macroscopic model describes the average dynamics of a polymerase/template complex based on the details of a kinetic pathway.

In Fig. 1 a kinetic scheme is presented that is based on the mechanism that Johnson (1993) proposed for DNA propagation by T7 polymerase. The template DNA is designated as D_n to indicate that the extension has progressed to position n and the insertion of the $n+1$ nucleotide is considered. The polymerase is E and the deoxynucleotide triphophate is dNTP. The dNTP pool comprises of dATP, dCTP, dGTP and dTTP. The state before the enzyme binds to the primer/template complex is denoted as *A1*. The complex binds a polymerase molecule to form a ternary complex *A0*. There is a small modification to the original scheme of Johnson, notably the detachment of the polymerase from the quaternary state, i.e., *A2→A1*. (Due to the paucity of kinetic data, it is reasonable to assume that k_5 and k_{-1} would be similar.)
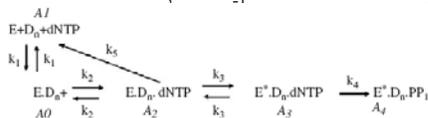

Fig. 1. Schematic of polymerase-catalyzed DNA elongation. *A1* presents the primer/template complex, *A0* is the ternary complex that consists of the template:primer:polymerase, *A2* is the quaternary complex that forms after binding of the ternary complex with a dNTP. The model accounts for polymerase detachment from states *A0* and *A2*. The quaternary complex adopts an activated state *A3* followed by phosphodiester bond synthesis, the release of pyrophosphate, and translocation of the enzyme. The schematic also identifies the notation to label the different states and the respective rate constants.

Let $[X]$ present the sum of all the species in the dNTP pool, then the molar fraction of each species is expressed as x_i (where $i = A, C, T, G, U$). After one of the components of the dNTP pool has bound to the ternary complex to form the quaternary complex *A2*, the polymerase may undergo a conformational change denoted as *A3*. The final step to *A4* involves the release of PP_i. Patel et al. (1991) noted that the rate-limiting step is the formation of the activated complex *A3*. The state *A4* is a ternary complex similar to *A0*, but the number of base pairs has increased by one. Viljoen et al. (2005) derived an expression for the average extension rate. It must be noted that the macroscopic model lends itself to a hierarchy of complexity. At the basic level, no distinction is made between the insertion times of different nucleotides. In a further refinement the GC contents of the template in the vicinity of the insertion site could be included and finally the effect of secondary structures of the template could be included. The average time to add a nucleotide to a template with composition $N = N_A + N_T + N_C + N_G$ is

$$
\begin{aligned}
t_{ave} &= \frac{1}{v_{ave}} = \frac{1}{N} \sum_{i=A,C,T,G} \frac{N_i[x_i\tau/P_S + (1-x_i)\tau_I/P_S]}{x_i + (1-x_i)P_{SI}/P_S} \\
&\approx \frac{\tau}{P_S} + \left(\sum_{i=A,C,T,G} \frac{N_i(1-x_i)}{Nx_i} \right) \frac{\tau_I}{P_S} \text{(s/nucleotide)}. \quad (1)
\end{aligned}
$$

The four parameters τ, τ_I, P_S, and P_{SI} are functions of the rate constants in Fig. 1. The physical interpretation of the four parameters τ, P_S, τ_I, and P_{SI} is as follows. The probability to insert a correct nucleotide is P_S and the probability to insert an incorrect nucleotide is $P_{SI} \ll 1$. The parameters τ and τ_I are the average passage times for a correct (τ) or an incorrect nucleotide (τ_I). The denominator in Eq. (1) contains the term P_{SI}/P_S and it is expected that $P_{SI}/P_S \ll 1$, because polymerases are efficient. It has been shown in the case of KOD Pol (Griep et al., 2006) that $P_{SI}/P_S \approx 10^{-6}$. Under most circumstances the denominator in Eq. (1) is approximately x_i — an exception occurs if an extreme dNTP pool composition $x_i \ll 1$ is used. Therefore it is argued that the approximate expression on the right hand side of Eq. (1) is a valid simplification. The kinetic model further provides an estimate of the number of errors which are made after extension of a DNA template of length N:

$$
E = \left\{ \sum_{i=A,C,T,G} \frac{N_i(1-x_i)\Phi}{x_i} \right\}. \quad (2)
$$

The parameter Φ in Eq. (2) is characteristic for a specific type of DNA polymerase. It can be estimated from published data. For example, Mizuguchi et al. (1999) reported the error frequency of KOD Pol under equimolar pool conditions as $\Phi = 1.1 \times 10^{-6}$ (i.e., 1.1 errors per million base pairs). Till date the P.I's are not aware of published information on the temperature dependence of Φ.

Eq. (1) appears deceptively simple. Consider the states *A0, A1… A4* in Fig. 1. There is a finite probability to proceed from any state to a connected (connected by a reaction) state. For example, the probability to proceed from state *A0* to state *A2* is

$$P_{02} = \frac{k_2[X]}{k_2[X] + k_{-1}}. \tag{3}$$

Note that the first integer of the subscript of P_{IJ} corresponds to the departing state and the second integer corresponds to the arrival state. The probability to extend the sequence by a correct nucleotide is

$$P_S = P_{02}P_{23}P_{34}/[1-P_{23}P_{32}]. \tag{4}$$

The remaining probabilities in Eq. (4) are defined as

$$P_{23} = \frac{k_3}{k_3 + k_{-2} + k_{-1}}; \tag{5a}$$

$$P_{32} = \frac{k_{-3}}{k_4 + k_{-3}}; \tag{5b}$$

$$P_{34} = 1 - P_{32} = \frac{k_4}{k_4 + k_{-3}}. \tag{5c}$$

The average passage time for a correct nucleotide is

$$\tau = P_S[t_{02}+t_{23}+t_{34}]+P_SP_{23}P_{32}[t_{23}+t_{32}]/(1-P_{23}P_{32}). \tag{6}$$

If an incorrect nucleotide binds to the ternary complex, the forward rate constants k_3 and k_4 change to lower values k_{3I} and k_{4I} (Fersht et al., 1983 and Florian et al., 2002). Note however that the nucleotides still bind to the ternary complex with the same rate constant k_2. The expressions for P_{SI} and τ_I correspond to Eqs. (4) (6), after replacing k_3 and k_4 with k_{3I} and k_{4I}. In comparison, the macroscopic kinetic model (1) contains only four parameters, which is a drastic reduction in complexity.

An important result of the analysis is the role that the dNTP pool plays. From biochemical and physiological viewpoints the imbalances of the intracellular dNTP pool are linked with enhanced rates of biological mutagenesis, cancer and genetic diseases. To begin with, dGTP is naturally underrepresented in eukaryotic cells and typically accounts for 5–10% of the total pool (that is, $x_G = 0.05$ to 0.10) (Mathews and Ji, 1992). Biochemical and cellular experiments show that eukaryotic polymerases are slightly but significantly more mutagenic in these naturally imbalanced dNTP pools than in equimolar pools (Martomo and Mathews, 2002). This is exactly the modest outcome predicted by our model. When the pool is biased beyond the natural, it should cause substantial increases in the mutagenic rates. In fact, a review of biological results (Viljoen et al., 2005) indicates that drastic changes to the dNTP pools do occur when cells are treated with many nucleoside analogs or when the nucleoside biosynthetic or salvaging pathways are mutated or altered in some way.

3. Experimental method

3.1. Approach

The approximate form on the right hand side of Eq. (1) contains only three parameters, but they are present in the form $\Gamma = \tau/P_S$ and $\Gamma_I = \tau_I/P_S$. The parameters (Γ, Γ_I) are unique for a specific type of polymerase and we posit the following dependence on temperature:

$$\tau/P_S = \Gamma = \Gamma_0 e^{-E/RT}; \tag{7a}$$

$$\tau_I/P_S = \Gamma_I = \Gamma_{I0} e^{-E_I/RT}. \tag{7b}$$

After Eq. (1) has been multiplied by the template length (excluding the primer length), the elongation time t_{El} is obtained;

$$Nt_{ave} = t_{El} \approx N\Gamma + \left(\sum_{i=A,C,T,G} \frac{N_i(1 - x_i)}{x_i} \right) \Gamma_I. \tag{8}$$

Accurate measurements of the elongation times at different temperatures and dNTP pool compositions are needed to set up a set of independent equations to solve for the unknown parameters $(\Gamma_0, E, \Gamma_{I0}, E_I)$. First a template with a known sequence is PCR amplified at an elongation temperature T_1 with an equimolar dNTP pool. The experiment is repeated with progressively shorter elongation times until the cut-off is reached—any reduction in elongation time below the cut-off value leads to a sharp drop in the intensity of the gel electrophoresis results. The experiments are repeated with a skewed dNTP pool at the same elongation temperature T_1. Next the elongation times are measured for an equimolar and a skew pool composition at a different elongation temperature T_2. The four elongation times at two different temperatures and two different pool compositions are fitted to an Arrhenius plot to solve for (Γ_0, E) and (Γ_{I0}, E_I).

3.2. PCR experiments

The amplicon that has been used in this study is a 52% GC-rich, 2517-basepair puc19 DNA fragment, from *E. coli*. The primers are 22 base pairs long and thus the effective length of the amplicon is $N = (2517-22)$ bp. The DNA polymerase is from KOD in its hot start formulation (Mizuguchi et al., 1999). The experiments were carried out in the PCRJet thermocycler (Quintanar and Nelson, 2002 and Moore, 2005).

The reaction conditions for each 20 µl reaction were:

Initial incubation of polymerase at 94 °C for 30 s, denaturing at 94 °C for 1 s, annealing at 61 °C for 4 s, elongation temperature ranges between 65 °C, 70 °C, 72 °C, and 75 °C with variable elongation times.

A 100-µl volume of PCR reaction mixture has been prepared from:

- fractional volumes of 1 unit/µl of KOD hot start polymerase,
- 10X KOD hot start polymerase buffer (Novagen, Wisconsin, Madison), final concentration of 1X,

- 10 X dye consisting of sucrose and cresol (Idaho Technologies, ID), final concentration of 1X,
- Bovine serum albumin (BSA, Sigma chemicals, St. Louis, MO), 3 µl added for a final concentration of 600µg/µl BSA.
- A final concentration of 4 mM MgSO$_4$, 16 µl of 25 mM added.
- 200 mM deoxynucleotide triphosphate (dNTP) for equimolar amounts, 600µM dATP, 67 µM dCTP, 67 µM dTTP, 67 µM dGTP, for non-equimolar amounts.
- 1:1 ratio of 25 pmol reverse primers and 25 pmol forward primers, at a final concentration of 0.375 pmols
- 100 pg of pUC19 DNA, 6 µl added to overall (600 pg total).

After the PCR mix has been amplified, it is subjected to gel electrophoresis.

3.3. Thermocycler

To obtain accurate measurements of the elongation times, the thermocycler must be fast. For that reason the DNA is amplified in the PCRJet® thermocycler (Quintanar and Nelson, 2002 and Moore, 2005). The device controls the flow of hot and cold gases at high velocities with electronic valves. Fast thermocycling is achieved by the flow of high velocity air through a reaction chamber that houses an array of cuvettes. The inlet air streams (one hot, one cold) are pre-conditioned and mixed to achieve the required temperature and flow rate in the reaction chamber. Heat transfer to and from the reaction cuvette depends on forced convection. The temperature protocol is accurate to within 1 °C and has a response time of 100 ms. (Temperature is measured with thermocouples and sources of error include calibration, electric potential-temperature conversion, time delays as well as a measuring error with usual normal distribution. The accuracy reported here reflects the accuracy of measurement, but even in finely calibrated thermocouples it is realistic to expect an absolute error of 1 °C) (Omega Handbook and Encyclopedia,). The instrument is capable to perform thirty cycles of 0 s at 90 °C and 0 s 57 °C in 42 s, or 1.4 s/cycle. No PCR can be performed at these cycling speeds but it demonstrates that the rate-limiting step is the enzyme-catalyzed DNA extension.

In Fig. 2 a PCRJet® temperature–time profile is shown for one of the protocols used in the experiments—8 s elongation. The time segments are marked on the plot and the values are listed in Table 1. With the exception of segment t3 (extension time) all the segments are the same for all experiments. The heating/cooling rate which can be achieved with the PCRJet® depends on the dimensions of the cuvettes. If Lightcycler capillaries (www.roche-applied-science.com) with a 20 µl capacity are used (as is the case for the experiments described here), a rate of 40ºC/s has been used. The total time of the PCR cycle shown in Fig. 2 is 14.66 s. The annealing time in this study is 4.00 s and the denaturing time is 1.00 s.

The elongation time is stepped in intervals of 1.00 s and gel electrophoresis results are compared. A drastic reduction in band brightness marks the transition from the shortest success-
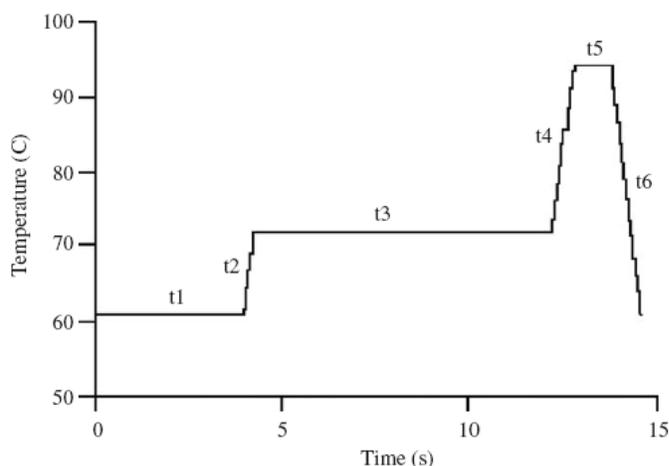

Fig. 2. Temperature–time plot for the PCR protocol used to amplify the 2517bp fragment of pUC19.

Table 1
Time segments of temperature–time protocol

	t1	t2	t3	t4	t5	t6	Total
Time (s)	4.00	0.28	8.00	0.55	1.00	0.83	14.66

ful elongation time (cut-off value) to values where elongation is unsuccessful. Since elongation starts at the end of period t_1, the total reaction time is t_2+t_3. (We neglect the ramp time from t_3.) In the analysis of reaction times we only consider t_3 as the reaction time, because the temperature changes during the ramp times. Therefore the correct reaction time is understated by:

$$f_{RX} = \frac{t_3}{t_2 + t_3} \quad \text{where} \quad t_2 = \frac{t_{el} - t_{an}}{40°C/s}.$$

The lowest value of f_{RX} is found for the balanced pool at 75 °C elongation. For the values in Table 1,

$$f_{RX} = \frac{9.00}{0.35 + 9.00} = 96.26\%.$$

4. Results

The experimental values of elongation times for the balanced and the biased dNTP pool compositions (cf. Section 3.2) are listed in Table 2. The data in Table 2 are used to calculate Γ and Γ_f values at all four temperatures and both dNTP pool compositions (Fig. 3).

4.1. Example of Γ and Γ_f calculation.

The composition of the template is

$$\left(\frac{N_A}{N}; \frac{N_C}{N}; \frac{N_T}{N}; \frac{N_G}{N}\right) = (0.239; 0.261; 0.239; 0.261)$$

and substitution of these values in Eq. (8) gives

$$t_{El} = N\Gamma + (0.239(1-x_A)/x_A + 0.261(1-x_C)/x_C$$
$$+ 0.239(1-x_T)/x_T + 0.261(1-x_G)/x_G) \times N\Gamma_f. \quad (9)$$

Table 2
Experimental values of elongation times

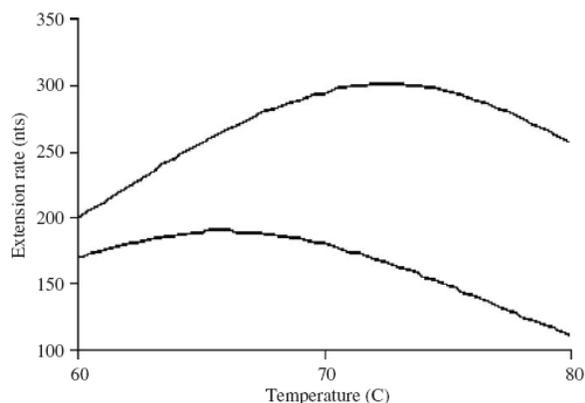| Temperature	Times for balanced pool	Times for biased pool	Γ (ms)	Γ_I (ms)
65	10, 11, 10	14, 14, 14	3.30	0.27
70	8, 8, 7	13, 12, 12	2.03	0.34
72	9, 8, 8	15, 15, 15	1.85	0.49
75	9, 9, 9	18, 17, 18	1.68	0.63

Fig. 3. Elongation rates (nt/s) for a balanced dNTP pool (upper line) and a biased dNTP pool (lower line).

Consider the extension times which have been measured at 65 °C. The average time for the balanced pool is $(11 + 11 + 10)s/3 = 10.33$ s, and the molar fractions of all dNTPs are 0.25. Substitute these values into Eq. (9);

$$10.33 = N\Gamma + 3 \times N\Gamma_I. \tag{10}$$

The average extension time for the biased pool is 14.00 s and the dNTP pool composition is $(x_A;x_C;x_T;x_G) = (600/800;67/800;67/800;67/800)$. When these values are substituted into Eq. (9) one obtains

$$14.00 = N\Gamma + 8.45 \times N\Gamma_I. \tag{11}$$

Eqs. (10), (11) are solved and the values of (Γ, Γ_I) at 65 °C are (3.30; 0.27). *Note: N = 2517.*

The values of Γ and Γ_I are listed in the last two columns of Table 2. Note that Γ decreases with temperature and Γ_I increases with temperature. The slopes and intercepts of the Arrhenius plots $\ln(\Gamma)$ vs. $1/T$ and $\ln(\Gamma_I)$ vs. $1/T$ give the activation energies and pre-exponent coefficients of Γ and Γ_I. Both plots have a correlation coefficient of 94%. Thus Γ and Γ_I can be expressed as

$$\Gamma = 1 \times 10^{-10} e^{16,131/RT}(ms), \tag{12}$$

$$\Gamma_I = 4 \times 10^{12} e^{-20,372/RT}(ms), \tag{13}$$

Note: R = 1.98 cal/mol K.

The pre-exponential factors and activation energies in Eqs. (12)–(13) are quite sensitive to changes in the experimental

data (Table 2). The problem is the multiplication of very large numbers with very small numbers. A similar problem is encountered in reaction engineering if the activation energy of the reaction is large (typically combustion reactions). Thus we follow the same approach as in reaction engineering; select 65 °C (i.e., T_{ex} = 338.15 K) as the expansion point and write Γ and Γ_I as follows;

$$\Gamma = [\Gamma_0 e^{E/RT_{ex}}]e^{-E_I(T-T_{ex})/RT_{ex}^2} = \Lambda e^{-\alpha(T-T_{ex})}, \tag{14}$$

$$\Gamma_I = [\Gamma_{I0}e^{-E_I/RT_{ex}}]e^{E_I(T-T_{ex})/RT_{ex}^2} = \Lambda_I e^{\alpha_I(T-T_{ex})}. \tag{15}$$

To investigate the sensitivity of the results, each of the four balanced pool data sets and the four unbalanced pool data sets in Table 2 have been varied by ± 0.33 s. (This choice is based on an error of ± 1 s in one of the data points of each set.) The variance and standard deviation have been calculated for each of the four parameters in Eqs. (14)–(15) and the final results are:

$$\Gamma = (2.95 \pm 0.20)e^{-(0.67 \pm 0.06)(T - T_{ex})}(ms), \tag{16}$$

$$\Gamma_I = (0.24 \pm 0.02)e^{(0.85 \pm 0.07)(T - T_{ex})}(ms). \tag{17}$$

Remark: The accuracy can be further improved by adding more data points, but it must be noted that each experiment involves a lot of work.

The gamma relations (16), (17) are used to predict the temperature-dependence of the DNA polymerase extension rates (Fig. 3). The lower curve corresponds to the biased dNTP pool and the top curve corresponds to the balanced pool. The maximum extension rate for KOD polymerase is 301 nts (nucleotides/s) and the maximum rate is reached at 72.5 °C. The extension rate is slowed down significantly when the dNTP pool is not balanced. In this case the maximum rate is 191 nts and it is reached at 66 °C — still slower than the balanced pool which has a value of 267 nts at the same temperature. Another interesting result is that the maximum extension rate is a function of pool composition. In retrospect it is to be expected, because the pool composition affects the role of the time-delaying and successful-insertion processes, all of which have different temperature dependencies. When the pool is unbalanced, the frequency of unsuccessful insertion attempts increases. Consequently Γ_I takes on a more dominant role—in Eq. (8) the coefficient of Γ_I changes from 3 for the balanced case to 8.45 for the unbalanced case. Since Γ_I is an increasing function of temperature, the optimum extension rate is shifted to a lower temperature.

In order to test the results (12), (13), we have first calculated the elongation time for the balanced pool at 63 °C from Eq. (6), using the data (12), (13). The elongation time is predicted to be 10.70 s. Next the experiments have been performed for a balanced pool at 63 °C for elongations times of 13, 12, 11 and 10 s. In Fig. 4 the gel electrophoresis results are shown. The bands are clearly visible for lane 1 (13 s) and lane 2 (12 s). No band is observed at 11 s (lane 3), which lies right at the cut-off value. The drastic change between 12 and 11 s is consistent with the theoretical prediction.

Fig. 4. Gel electrophoretic analysis of PCR products extended at 63 °C for various elongation times in the presence of a balanced dNTP pool: the MW ladder, 13 s elongation time (lane 1). 12 s elongation time (lane 2), 11 s elongation time (lane 3) and 10 s elongation time (lane 4).

5. Conclusions

The extension rates of polymerases depend on the composition of the dNTP pool and temperature. An expression has been derived to describe these dependencies on the average extension rate. The model contains four parameters and four linearly independent equations are obtained from measurements of the extension time at two different temperatures and two different pool compositions.

The main results of the study are:

1. Kinetic information of the polymerase can be determined with a fast thermocycler. The accuracy of extension times increases with faster heating rates and longer templates. The device that has been used in this study can heat at an average rate of 40ºC/s. For the experiments described in this study, the ramp time from the annealing temperature to the extension temperature never exceeds 4% of the extension time.

2. The parameter Γ is a measure of the average processing time (passage time) for a correct nucleotide. Actually Γ is an upper limit of the true time τ, since τ is divided by the probability P_S, which is just smaller than one. The value of Γ changes from 3.30 ms at 65 °C to 1.68 ms at 75 °C. Therefore the polymerase becomes more efficient to insert a nucleotide at higher temperature.

3. The parameter Γ_I is a measure of the average time it takes the polymerase to process an incorrect nucleotide. The processing of an incorrect nucleotide has two outcomes, on the rare occasion it incorporates the nucleotide, but in most cases the nucleotide is rejected. Γ_I is less than Γ, therefore one concludes that the polymerase processes an incorrect nucleotide faster at this temperature. Clearly the translocation step adds a significant amount of time to the successful insertion event. Γ_I increases with temperature. Included in Γ_I is the time-consuming step of enzyme detachment/re-attachment and it is known that enzyme detachment increases with temperature.

4. The optimum temperature of KOD polymerase activity depends on the dNTP pool. In the case of an equimolar composition, the optimum temperature is 72.5 °C. When the pool is biased to be $x_A = 0.75$; $x_{C, T, G} = 0.08375$, the optimum temperature drops to 66 °C. The general trend is that the optimum temperature is lower when the pool is not balanced.

5. The optimum temperature of KOD polymerase activity results from the competition between increasing amount of time spent in wasteful loops as temperature rises and increased activity with temperature. This is predicted theoretically from the model and shown experimentally. If the thermal optimum were due to the equilibrium between an active and inactive conformation, the thermal optimum would not shift as a result of dNTP pool composition. Therefore, the conformation model for thermal optima does not apply in the case of thermophilic DNA polymerases.

6. The macroscopic model provides the necessary tools for quantitative modeling of PCR (Peterson et al., 2004) and the design of experiments, specifically the length of the extension step and the temperature of extension. The optimization of reaction conditions takes on a new importance whenever the error frequency must be minimized. In that case, it becomes even more important to minimize the exposure time to high temperature. Although errors do not play a critical role in most diagnostic applications, there are applications where high fidelity replications are required.

7. The optimum of the elongation rate vs. temperature is flatter than expected. A local optimum is always indicative of opposing factors.

References

Ahern, T.J., and Klibanov, A.M. Analysis of processes causing thermal inactivation of enzymes, *Methods of Biochemical Analysis* **33** (1988), pp. 91–127.

Ahmed, A., and Tollefsbol, T.O. Telomerase, telomerase inhibition, and cancer, *Journal of Anti Aging Medicine* **6** (2003), pp. 315–325.

von Ahsen, N., Wittwer, C.T., and Schutz, E. Oligonucleotide melting temperatures under PCR conditions: nearest–neighbor corrections for Mg(2+), deoxynucleotide triphosphate, and dimethyl sulfoxide concentrations with comparison to alternative empirical formulas, *Clinical Chemistry* **47** (2001), pp. 1956–1961.

Arezi, B., Xing, W., Sorge, J.A., and Hogrefe, H.H. Amplification efficiency of thermostable DNA polymerases, *Analytical Biochemistry* **321** (2003), pp. 226–235.

Benita, Y., Oosting, R.S., Lok, M.C., Wise, M.J., and Humphery-Smith, I. Regionalized GC content of template DNA as a predictor of PCR success, *Nucleic Acids Research* **31** (2003), p. e99.

Daniel, R.M., Danson, M.J., and Eisenthal, R. The temperature optima of enzymes: a new perspective on an old phenomenon, *Trends in Biochemical Sciences* **26** (2001), pp. 223–225.

Datta, K., and LiCata, V.J. Thermodynamics of the binding of *Thermus aquaticus* DNA polymerase to primed–template DNA, *Nucleic Acids Research* **31** (2003), pp. 5590–5597.

A.R. Fersht, J.P. Shi and W.C. Tsui, Kinetics of base misinsertion by DNA polymerase I of *Escherichia coli*, *Journal of Molecular Biology* **165** (1983), pp. 655–667.

Florian, J., Goodman, M.F., and Warshel, A. Theoretical investigation of the binding free energies and key substrate-recognition components of the replication fidelity of human DNA polymerase beta, *Journal of Physical Chemistry B* **106** (2002), pp. 5739–5753

Griep, M., Whitney, S.E., Nelson, R.M., and Viljoen, H.J. DNA polymerase chain reaction: prediction of error frequencies and extension rates, *A.I.Ch.E. Journal* **52** (2006), pp. 384–392.

Grundberg-Manago, M., Oritz, P.J., and Ochoa, S. Enzymatic synthesis of nucleic acid-like polynucleotides, *Science* **122** (1955), pp. 907–910.

Innis, M.A., Myambo, K.B., Gelfand, D.H., and Brow, M.A. DNA sequencing with *Thermus aquaticus* DNA polymerase and direct sequencing of polymerase chain reaction—amplified DNA, *Proceedings of the National Academy of Science USA* **85** (1988), pp. 9436–9440.

Johnson, K.A. Conformational coupling in DNA polymerase fidelity, *Annual Review Biochemistry* **62** (1993), pp. 685–713.

Karantzeni, I., Ruiz, C., Liu, C.C. and LiCata, V.J. Comparative thermal denaturation of *Thermus aquaticus* and *Escherichia coli* Type 1 DNA polymerases, *Biochemistry Journal* **374** (2003), pp. 785–792.

Klibanov, A.M., and Ahern, T.J. Thermal stability of proteins. In: D.L. Oxender and C.F. Fox, Editors, *Protein Engineering*, Alan R. Liss, New York (1987), pp. 213–218.

Lehman, I.R. Discovery of DNA polymerase, *Journal of Biological Chemistry* **278** (2003), pp. 34, 733–34, 738.

Lumry, R., and Eyring, H. Conformation changes of proteins, *Journal of Physical Chemistry* **58** (1954), pp. 110–120.

Martomo, S.A., and Mathews, C.K. Effects of biological DNA precursor pool asymmetry upon accuracy of DNA replication in vitro, *Mutation Research* **499** (2002), pp. 197–211.

Mathews, C.K., and Ji, J. DNA precursor asymmetries, replication fidelity, and variable genome evolution, *Bioessays* **14** (1992), pp. 295–301.

Miura, S., and Izuta, S. DNA polymerases as targets of anticancer nucleosides, *Current Drug Targets* **5** (2004), pp. 191–195.

Mizuguchi, H., Nakatsuji, M., Fujiwara, S., Takagi, M., and Imanaka, T. Characterization and application to hot start PCR of neutralizing monoclonal antibodies against KOD DNA polymerase, *Journal of Biochemistry* **126** (1999), pp. 762–768.

Moore, P. PCR: Replicating success, *Nature* **435** (2005), pp. 235–238.

Ninio, J. Alternative to the steady-state method: derivation of reaction rates from first-passage times and pathway probabilities, *Proceedings of the National Academy of Science USA* **84** (1987), pp. 663–667.

Omega Handbook and Encyclopedia, vol. 1.

Owczarzy, R., *et al.*, Effects of sodium ions on DNA duplex oligomers: improved predictions of melting temperatures, *Biochemistry* **43** (2004), pp. 3537–3554.

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

Peterson, M.E., Eisenthal, R., Danson, M.J., Spence, A., and Daniel, R.M. A new intrinsic thermal parameter for enzymes reveals true temperature optima, *Journal of Biological Chemistry* **279** (2004), pp. 20, 717–20, 722.

Quintanar, A., Nelson, R.M., 2002. A process and apparatus for high–speed amplification of DNA. US Patent # 6, 472, 186.

Saiki, R.K., *et al.* Enzymatic amplification of beta—globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia, *Science* **230** (1985).

Saiki, R.K., *et al.* Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase, *Science* **239** (1988), pp. 487–491.

Smith, A.J. DNA sequence analysis by primed synthesis, *Methods in Enzymology* **65** (1980), pp. 560–580.

Viljoen, S., Griep, M.A., Nelson, M., and Viljoen, H. A macroscopic kinetic model for DNA polymerase elongation and high-fidelity nucleotide selection, *Computational Biology and Chemistry* **29** (2005), pp. 101–110.

Whitney, S.E., Sudhir, A., Nelson, R.M., and Viljoen, H.J. Principles of rapid polymerase chain reactions: mathematical modeling and experimental verification, *Computational Biology and Chemistry* **28** (2004), pp. 195–209.