Available online at www.sciencedirect.com
SCIENCE DIRECT®
Computational Biology and Chemistry
ELSEVIER
Computational Biology and Chemistry xxx (2006) xxx–xxx
www.elsevier.com/locate/compbiolchem

A quantitative model of error accumulation during PCR amplification

E. Pienaar [a], M. Theron [b], M. Nelson [c], H.J. Viljoen [a,*]

[a] *Department of Chemical Engineering, University of Nebraska, Lincoln, NE 68588, USA*
[b] *Department of Human Genetics, University of the Free State, NHLS, Bloemfontein 9330, South Africa*
[c] *Megabase Research Products, Lincoln, NE 68504, USA*

Received 14 November 2005; accepted 20 November 2005

Abstract

The amplification of target DNA by the polymerase chain reaction (PCR) produces copies which may contain errors. Two sources of errors are associated with the PCR process: (1) editing errors that occur during DNA polymerase-catalyzed enzymatic copying and (2) errors due to DNA thermal damage. In this study a quantitative model of error frequencies is proposed and the role of reaction conditions is investigated. The errors which are ascribed to the polymerase depend on the efficiency of its editing function as well as the reaction conditions; specifically the temperature and the dNTP pool composition. Thermally induced errors stem mostly from three sources: A + G depurination, oxidative damage of guanine to 8-oxoG and cytosine deamination to uracil. The post-PCR modifications of sequences are primarily due to exposure of nucleic acids to elevated temperatures, especially if the DNA is in a single-stranded form. The proposed quantitative model predicts the accumulation of errors over the course of a PCR cycle. Thermal damage contributes significantly to the total errors; therefore consideration must be given to thermal management of the PCR process.

Keywords: Polymerase chain reaction; Thermal damage; Depurination; Cytosine deamination

1. Introduction

The use of PCR to amplify a DNA target and cloning of single copy genes has become an important process in molecular biology. In most applications the objective is to identify the presence or absence of a specific DNA fragment and the error frequency is not of utmost importance. However, errors play a role in polymerase stability and in some instances the presence of an error could terminate the extension process. As the interest in synthetic DNA grows, techniques like PCA (polymerase chain assembly, Stemmer et al., 1995) are used to synthesize DNA molecules. One aspect of DNA synthesis that is critical is the control of errors. Therefore, a quantitative model of the error frequencies of DNA products from PCR may be helpful to adjust reaction conditions, reaction compositions, and the choice of polymerase.

One source of error stems from editing mistakes of the enzyme when the annealed oligomers are enzymatically extended. Extension errors can be reduced with a high-fidelity polymerase enzyme and optimization of the biochemical reaction conditions during DNA extension. *Pfu* polymerase (*Pyrococcus furiosus*) has outstanding fidelity, but its extension rate is very slow (∼20 nt/s at 72 °C). *Thermus aquaticus* polymerase (*Taq Pol*) has an elongation rate of approximately 80 nt/s at 72 °C (Innis et al., 1988; Gelfand and White, 1990; Whitney et al., 2004), but has no 3′ editing activity. Faster thermostable β-type DNA polymerases such as *Pyrococcus kodakaraensis* (*KOD Pol*; Toyobo Co. Ltd., Osaka; Mizuguchi et al., 1999) or *KlenTaq*22 (Cline et al., 1996) are preferred for high-speed amplification of DNA. In practice, the error rate of *KOD Pol* is extremely low during conditions of high-speed PCR. Mizuguchi et al. (1999) reported an error rate of ∼1.1 errors/10^6 bp for *KOD Pol*.

A major contributor to the errors in synthetic DNA molecules is thermally induced damage. Errors due to thermal damage stem mostly from three sources: A + G depurination (Lindahl and Nyberg, 1972), oxidative damage of guanine to 8-oxoG (Cadet et al., 2002; Hsu et al., 2004) and cytosine deamination to uracil (Lindahl and Nyberg, 1974). Depurination involves the removal

* Corresponding author at: Department of Chemical Engineering, 211 Othmer Hall, Lincoln, NE 68588-0643, USA. Tel.: +1 402 472 9318; fax: +1 402 472 6989.
E-mail address: hviljoen1@unl.edu (H.J. Viljoen).

1476-9271/$ – see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.compbiolchem.2005.11.002

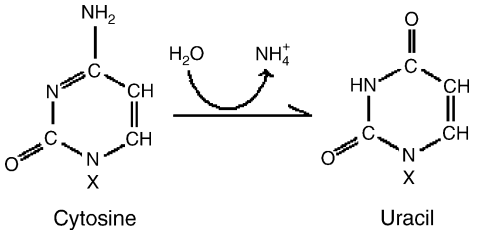

Fig. 1. Schematic representation of cytosine deamination to uracil. Here the letter X represents the ribose phosphate part of the nucleotide.

of a purine (adenine or guanine) from the DNA molecule, leaving only the sugar and phosphate backbone (attached to the DNA molecule). Cytosine deamination is illustrated in Fig. 1. These modifications of sequences are primarily due to exposure of nucleic acids to elevated temperatures, especially if the DNA is in a single-stranded form. Thermal damage results in either incorrect nucleotides being inserted into the complementary strand or the polymerase stalling at the abasic site. Thus consideration must be given to reduce damage to DNA in the rational design of the process. For example, oxidative damage can be reduced in PCR experiments by purging mixtures with argon to remove dissolved oxygen.

The effects of thermal damage are also reflected in the product yield. Degradation of the template DNA (especially during pre-amplification denaturation) has been correlated with a significant decrease in product yield (Gustafson et al., 1993, Sikorsky et al., 2004). Yap and Mcgee (1991) suggested lowering the denaturing temperatures after the first few cycles in order to improve product yield.

Few precautions are taken to minimize DNA thermal damage during PCR experiments and gene assembly experiments. For example, Smith et al. (2003) utilized a 16 h *Taq* ligase reaction at 65 °C. Based on the published rate constants for A + G depurination and C deamination, one expects very high levels of errors when such overnight heating of DNA is carried out. In particular, the rate constants published by Fryxell and Zuckerkandl (2000) and Lindahl and Nyberg (1972, 1974) predict that A + G depurination and C deamination will reach levels of 0.2–0.3% after 1 h at 72 °C; i.e. one in every 300–500 bases will be thermally damaged.

Furthermore, the thermocyclers used in these experiments are slow and have been used in combination with conservatively long thermocycling protocols. For example, in the widely used PCR protocol of Saiki et al. (1985), 1 min temperature holds at each of three temperatures 94, 55, and 72 °C are employed. Therefore DNA unnecessarily spends 2 min/cycle at temperatures greater than 70 °C, during which time A + G depurination and C deamination occur.

It is concluded from the discussion that high-fidelity synthesis of DNA molecules requires an awareness of biosynthetic errors and a practical strategy to minimize these errors. Error control requires a fundamental understanding of (1) the relationship between the kinetics of the polymerase chain reaction and errors which occur during DNA replication and (2) thermal management of the process to minimize high temperature exposure. In summary, the combination of a fast thermocycler, in which DNA spends very little time at elevated temperature and kinetically optimized DNA biochemistry, is the optimum strategy.

2. Mathematical model of error formation

The two sources of errors which occur during PCR amplification of DNA are (1) mistakes made by the polymerase and (2) thermal damage of the DNA in double- and single-stranded form. If mistakes occur early in the PCR process, the erroneous templates could lead to a large number of DNA copies with mistakes in their sequence. On the other hand, if mistakes occur later in the PCR process and/or incorrect templates form a small fraction of the total template population, then the total error concentration will be low. Furthermore, if a DNA polymerase is used that does not process certain errors (such as uracil insertion or C deamination), then these erroneous templates are not amplified. We need to distinguish between (1) how errors originate and (2) how errors proliferate. In this study the focus is on the origin of errors, therefore the mathematical model only describes the errors, which occur when an authentic template is copied during one PCR cycle.

Modeling strategy. The modeling strategy is summarized as follows.

(1) A template is selected and the PCR conditions are specified.
(2) The PCR cycle is divided into N segments of 10 ms each; s_1, s_2, . . ., s_N. The segment numbering starts at the beginning of the annealing step. A temperature is assigned to each segment; $T(s_1)$, $T(s_2)$, . . ., $T(s_N)$.
(3) The number of nucleotides which are added to the template over the interval s_j is calculated from the polymerase kinetic model, described in Section 2.1.
(4) For temperature $T(s_j)$, the degree of melting is calculated for the double stranded part of the template by the Monte Carlo method described in Section 2.3.
(5) The number of ds[A + G], ss[A + G], ds[C] and ss[C] are calculated.
(6) The rates of double (single) stranded depurination and C deamination are calculated at $T(s_j)$. The rate expressions are described in Section 2.2. After the respective rates are multiplied by the time interval s_j, the contributions are added to the respective cumulative errors.
(7) The time is advanced to s_{j+1} and steps (3)–(6) are repeated.
(8) At the end of segment s_N the errors produced by the polymerase catalyzed extension (c.f. Section 2.1) are added to the cumulative thermal errors.

Remark. We distinguish between the rates for the template and the complementary strand. A four letter notation is used to identify rates: the first two letters indicate single or double stranded kinetics (‘ss’ or ‘ds’), the third letter identifies the type of reaction—‘d’ is depurination, ‘c’ is C deamination. The last letter identifies the strand, ‘t’ is the template and ‘c’ is the complementary strand.

2.1. Kinetics and error production by polymerase

A better understanding of the fidelity of the polymerase chain reaction is obtained from a perspective of the polymerase kinetics. For that reason, a brief discussion on the kinetics of the polymerase chain reaction is given. The kinetic mechanisms and structure/function relationships of T7 DNA polymerase, *Klen-Taq* DNA polymerase, and a *Bacillus* DNA polymerase which have been worked out in some detail (Johnson, 1993; Kiefer et al., 1998; Li and Waksman, 2001; Patel et al., 1991), suggested that all DNA polymerases have the same broadly conserved molecular properties and mechanisms, but that each polymerase also has unique features (Jager and Pata, 1999).

Viljoen et al. (2005) and Griep et al. (2006) developed a macroscopic model of PCR kinetics which is based on the probabilistic kinetic approach described by Ninio (1987). The principle idea of Ninio's approach is to track a single enzyme/template complex over time and to determine its average behavior. The main results of the analysis macrokinetics model are expressions for the average extension rate v_{ave} (or its inverse t_{ave}—the average insertion time per nucleotide) and the error frequency.

The extension rate depends explicitly on template composition N_i (i = A, C, T, G), the dNTP pool composition, expressed as molar fractions, e.g. x_A = [dATP]/[dNTP] and it depends implicitly on the temperature via the model parameters.

2.1.1. Average extension rate

The average time to insert a nucleotide is:

$$t_{ave}\,(\mathrm{s/nt}) = \frac{1}{v_{ave}} = \frac{1}{N}\sum_{i=\mathrm{A,C,T,G}} \frac{N_i[x_i\tau/P_S + (1-x_i)\tau_I/P_S]}{x_i + (1-x_i)P_{SI}/P_S} \tag{1}$$

The reaction pathway, as discussed by Patel et al. (1991), involves several steps. Forward and reverse rate constants are associated with each step. The parameters τ, P_S, τ_I, and P_{SI} in Eq. (1) are functions of these rate constants (see Viljoen et al., 2005 for details). The parameters τ, P_S, τ_I, and P_{SI} can be viewed as functions of the kinetic rate constants of the reaction pathway, but a more tangible interpretation of the parameters is as follows. The probability to insert a correct nucleotide is P_S and the probability to insert an incorrect nucleotide is $P_{SI} \ll 1$ (because polymerases are efficient). The parameters τ and τ_I are the average passage times for a correct (τ) or an incorrect nucleotide (τ_I). The right hand side of Eq. (1) contains only three parameters; $\Gamma = (\tau/P_S)$, $\Gamma_I = (\tau_I/P_{SI})$ and $\Phi = P_{SI}/P_S$. The parameters (Γ, Γ_I, Φ) are functions of the type of polymerase and temperature. (Since the parameters are derived from kinetic constants, we posit Arrhenius forms of temperature dependence.) Other factors such as the (G + C) contents of the template and its size will also play a role.

The brief discussion on the kinetics helps to show the connection between the kinetics and errors. Eq. (1) is also used to calculate the progression of the DNA extension during a PCR cycle. This information is needed to determine the fraction of total template length that is single stranded.

2.1.2. Error frequency of polymerase extension

The kinetic model (1) provides an estimate of the number of errors which are made after extension of a DNA template of length N:

$$E = \left\{\sum_{i=\mathrm{A,C,T,G}} \frac{N_i(1-x_i)P_{SI}}{x_iP_S + (1-x_i)P_{SI}}\right\}/N \approx \left\{\sum_{i=\mathrm{A,C,T,G}} \frac{N_i(1-x_i)\Phi}{x_i}\right\} \tag{2}$$

The parameter $\Phi = P_{SI}/P_S$ in Eq. (2) is characteristic for a specific type of DNA polymerase and can be estimated from published data. For example, Mizuguchi et al. (1999) reported the error frequency of *Pyrococcus kodakaraensis* (*KOD Pol*) under equimolar dNTP pool conditions as $\Phi = 1.1 \times 10^{-6}$ (i.e. 1.1 errors/million base pairs). Till date the authors are not aware of published information on the temperature dependence of Φ, but such a relation certainly exists. Therefore the error frequency model in Eq. (2) will be used in the overall error model without accounting for changes in temperature during a PCR cycle.

It follows from Eq. (2) that the dNTP pool composition influences the error frequency. If the template consists of approximately equal amounts of A, C, T and G, the optimum dNTP pool composition is $x_A = x_C = x_G = x_T = 0.25$. If the dNTP pool is skew, the polymerase has to process an increased number of incorrect nucleotides and the extension rate slows down. If the template has a high G + C content, the optimum dNTP pool composition shifts towards higher molar fractions of C and G. Thus an optimum dNTP pool composition is associated with a template. The contrapuntal results are that an imbalanced dNTP pool increases t_{ave} and increases the error frequency. The following important conclusion came from the study by Griep et al. (2006): The error frequency is a minimum at the optimum dNTP pool composition.

2.1.3. Kinetic parameter values

The parameters (Γ, Γ_I, Φ), which are used in this study pertain to the β-type polymerase *Pyrococcus kodakaraensis* (*KOD Pol*). Mizuguchi et al. (1999) reported the error frequency of *Pyrococcus kodakaraensis* (*KOD Pol*) under equimolar dNTP pool conditions as $\Phi = 1.1 \times 10^{-6}$. The parameters Γ and Γ_I are assumed to have Arrhenius forms of temperature dependence (since they are functions of rate constants). For a selected amplicon, the minimum elongation times have been measured at different elongation temperatures and dNTP pool compositions. The amplicon that has been used in the study is a 52% GC-rich, 2517-bp fragment of pUC19 (Griep et al. (2005)). The minimum elongation times have been measured at different elongation temperatures and different dNTP pool compositions using a PCRJet® thermocycler (Megabase Research Products, Lincoln, NE). The results are as follow:

$$\Gamma\,(\mathrm{ms}) = 3.5 \times 10^{-12} e^{18.5/RT} \tag{3}$$

$$\Gamma_I\,(\mathrm{ms}) = 4 \times 10^{13} e^{-22/RT} \tag{4}$$

where $R = 0.00198$ kcal/mol K

Note that Γ and Γ_I have opposing temperature dependence. The parameter Γ describes the characteristic time to insert a correct nucleotide and it becomes shorter if the temperature is increased. In contrast, the parameter Γ_I increases with temperature due to lower processivity. The parameters Γ and Γ_I are calculated at a given temperature $T(s_j)$ and used in Eq. (1) to determine the number of nucleotides which are added during interval s_j.

2.2. Thermal damage

The mathematical model of thermal damage focuses on two reactions; (1) cytosine deamination and (2) depurination. Cytosine deamination is slowest at pH 8–8.5, the rate increases sharply at higher pH and more gradually at lower pH and it exhibits a small positive salt effect. The deamination rates are different for ssDNA and dsDNA. Fryxell and Zuckerkandl (2000) used data from three different sources to obtain the following rate expression for dsDNA:

C deamination of double-stranded DNA

$$k_{\mathrm{Cds}}\,(\mathrm{s}^{-1}) = 2.66 \times 10^{10} \mathrm{e}^{-32/RT},$$
$$(R = 0.00198\,\mathrm{kcal/mol\,K}) \quad (5)$$

Lindahl and Nyberg (1974) measured deamination rates of cytosine in denatured *E. coli* DNA. At 95 and 80 °C the rates are 2.2×10^{-7} and $1.3 \times 10^{-8}\,\mathrm{s}^{-1}$, respectively. If an Arrhenius plot is fitted between the two points, the rate constant is;

C deamination of single-stranded DNA

$$k_{\mathrm{Css}}\,(\mathrm{s}^{-1}) = 1.8 \times 10^{22} \mathrm{e}^{-48.6/RT} \quad (6)$$

Cytosine deamination of ssDNA does not only have a much larger activation energy than the deamination rate of dsDNA, but the reaction is much faster over the temperature ranges 60–95 °C.

Depurination reactions are strongly catalyzed at low pH. Based on the experimental data of Lindahl and Nyberg (1972), the rate constant for depurination of dsDNA at pH 7.4 is;

A + G depurination of double-stranded DNA

$$k_{\mathrm{DPds}}\,(\mathrm{s}^{-1}) = 2.3 \times 10^{11} \mathrm{e}^{-31/RT} \quad (7)$$

The depurination rate constant of ssDNA is nominally 3.3 times higher than for dsDNA (c.f. Fig. 8 in Lindahl and Nyberg, 1972);

A + G depurination of single-stranded DNA

$$r_{\mathrm{DPss}}\,(\mathrm{s}^{-1}) = 7.6 \times 10^{11} \mathrm{e}^{-31/RT} \quad (8)$$

The total accumulation of errors during PCA consists of the sum of thermal damage and PCR-based errors (c.f. Eq. (2)). To apply Eqs. (5)–(8) to a target template, it is necessary to know the degree of melting. At the beginning of a PCR cycle, defined as the completion of primer/template annealing, the bulk of the template is in single-stranded form. As extension progresses, the fraction of the template that is double stranded increases. However, it is not correct to assume that the polymerase/template complex marks the transition between dsDNA and ssDNA. Double-stranded DNA may form bubbles and localized melting may occur even at moderate temperatures. As the temperature increases to the denaturing temperature, the fraction of ssDNA template increases. Thus a model that describes the helix/coil transitions is needed to assess the thermal damage.

2.3. A model of dsDNA/ssDNA transition

The model that is proposed in this study is based on the assumption that the free energy of the DNA molecule depends on nearest neighbors (NN) and the salt concentration of the solution. The nearest neighbor concept was pioneered by Crothers and Zimm (1964), Gray and Tinoco (1970) and Uhlenbeck et al. (1973). Several experimental studies measured nearest neighbor interactions, specifically free energies and entropies. Santalucia (1998) provided a unified view of seven different studies and showed that six of the studies are in good agreement. The results are listed in Table 1. There has also been some controversy about the role of the length of the DNA molecule on nearest neighbor thermodynamics. Santalucia (1998) addressed the controversy and showed that there is a length dependence on the salt concentration, but no dependence exists for nearest neighbor energies. The values in Table 1 are free energy, measured in kcal/mol at 37 °C. When a pair is presented as AT/TA it means the 5′–3′ AT is paired with the 3′–5′ TA.

The last two columns list the values if the dsDNA molecule begins/ends with a GC or AT pair. For example, the oligomer AATGCC (5′–3′) has free energy

$$\Delta G_{37} = (-1.00) + (-0.88) + (-1.45) + (-2.24) + (-1.84) + (1.03) + (0.98) = -5.40\,\mathrm{kcal/mol}$$

The values in Table 1 are denoted as ΔG_{37}^{IK} (I, K = A, C, T, G). Usually the enthalpy is constant, but the free energy changes with temperature as follows:

$$\Delta G_T^{IK} = \Delta H^{IK} - T\Delta S^{IK} = \Delta G_{37}^{IK} - (T - 310.15)\Delta S^{IK} \quad (9)$$

In Table 2 the entropy is listed for all 10 different nearest neighbor combinations.

Santalucia (1998) proposed a salt correction for ΔS^{IK} and ΔG_{37}^{IK}, but these corrections are length dependent and results for oligomers shorter than 26 bp have been reported. The free energy of nucleation is corrected in similar manner, using the

Table 1
Free energy values of NN (kcal/mol)

AA/TT	−1.00
AT/TA	−0.88
TA/AT	−0.58
CA/GT	−1.45
GT/CA	−1.44
CT/GA	−1.28
GA/CT	−1.30
CG/GC	−2.17
GC/CG	−2.24
GG/CC	−1.84
Initial with GC	0.98
Initial with AT	1.03

Table 2
Entropy values of NN (cal/mol K)

AA/TT	−22.2
AT/TA	−20.4
TA/AT	−21.3
CA/GT	−22.7
GT/CA	−22.4
CT/GA	−21.0
GA/CT	−22.2
CG/GC	−27.2
GC/CG	−24.4
GG/CC	−19.9
Initial with GC	−2.8
Initial with AT	4.1

last two rows in Tables 1 and 2 in the right hand side of Eq. (10):

$$\Delta G_T^{\text{Init:AT,GC}} = \Delta G_{37}^{\text{Init:AT,GC}} - (T - 310.15)\Delta S^{\text{Init:AT,GC}} \tag{10}$$

The free energy due to nucleation depends on the type of base pair at position 1 and position N (first and last positions). After the temperature correction has been made as shown in Eq. (10), we denote this contribution to the free energy as $\Delta G_T^{\text{Nucl.}}$.

Two sources contribute directly to the free energy of a base pair; the hydrogen bond of the base pair and the stacking interaction. If the two contributions are combined, the doublet format is used. The singlet format refers to the separation of the two contributions. The data presented in Table 1 is in doublet format. Frank-Kamenetskii (1971) investigated the strength of the hydrogen bonds as a function of the salt concentration (molar) and expressed his findings in terms of temperature as follows:

$$T_{\text{AT}} = 355.55 + 7.95 \ln[\text{Na}^+] \tag{11a}$$

$$T_{\text{GC}} = 391.55 + 4.89 \ln[\text{Na}^+] \tag{11b}$$

The free energy associated with the hydrogen bond is

$$\Delta G_T^{\text{H:AT}} = (T_{\text{AT}} - T)\Delta S_{\text{H}}^{\text{AT}} \tag{12a}$$

$$\Delta G_T^{\text{H:GC}} = (T_{\text{GC}} - T)\Delta S_{\text{H}}^{\text{GC}} \tag{12b}$$

The entropy values are $\Delta S_{\text{H}}^{\text{AT}} = \Delta S_{\text{H}}^{\text{GC}} = -0.0224\,\text{cal/mol K}$ (c.f. Santalucia, 1998).

2.3.1. Monte Carlo method

A Monte Carlo method is used to determine the degree of melting. First the free energy associated with each base pair is determined. Next the free energy associated with the alternative state (helical to coil or vice versa) is calculated. A Metropolis decision-making algorithm is applied to each base pair based on the difference in free energy. The Monte Carlo procedure is repeated until steady state is reached. Consider the jth base pair in a template of length N. Define a state function S_j for each base pair, $S_j = 0$ if the jth base pair has a hydrogen bond, otherwise $S_j = 1$. Suppose the 5′–3′ composition at positions $j-1, j, j+1$ are 'ACG'. The nearest neighbor contributions to the free energy of the jth base pair is:

$$\Delta G_j^{\text{NN}} = \frac{1}{2}\left[\Delta G_T^{\text{GT}} - S_{j-1}\Delta G_T^{\text{H:AT}} - S_j \Delta G_T^{\text{H:GC}}\right] + \frac{1}{2}\left[\Delta G_T^{\text{CG}} - S_{j+1}\Delta G_T^{\text{H:GC}} - S_j \Delta G_T^{\text{H:GC}}\right] \tag{13}$$

The free energy due to nucleation $\Delta G_T^{\text{Nucl.}}$ is distributed amongst all base pairs with hydrogen bonds and the contribution to the jth base pair is:

$$\Delta G_j^{\text{Nucl.}} = \frac{\Delta G_T^{\text{Nucl.}}}{\sum_{k=1}(1 - S_k)} \tag{14}$$

The free energy of the jth base pair in its current state is therefore:

$$\Delta G_j = \Delta G_j^{\text{NN}} + \Delta G_j^{\text{Nucl.}} \tag{15}$$

The free energy of the jth base pair is calculated once again for the opposite state of base pair j. The state functions of all other base pairs are kept as before. Denote the free energy of the jth base pair in its opposite state as ΔG_j^{O}. Thus the contribution due to nucleation appears in only one of the two cases. Also note that for the end base pairs only one of the two terms in Eq. (10) is used.

After the energy difference ($\Delta E_j = \Delta G_j^{\text{O}} - \Delta G_j$) has been determined for all N base pairs, the probability for transition is calculated. The Metropolis algorithm defines the following transition probabilities:

$$P(S_j \to 1 - S_j) = \begin{cases} 1, & \Delta E_j \le 0 \\ \mathrm{e}^{-\Delta E_j/RT}, & \Delta E_j > 0 \end{cases} \tag{16}$$

To demonstrate the Monte Carlo method, it is applied to a sequence of 100 bp at 0.2 M Na^+ with varying G + C concentration. The melting curve is shown in Fig. 2 in comparison to the Marmur–Doty relation (c.f. Grisham and Garrett, p. 371–372 (1998)) that correlates melting temperature and G + C fraction: $T_{\text{M}} = 69.3 + 41 f_{\text{GC}}$. The maximum difference at the low G + C fraction is 2.5 °C and it is 2.7 °C at the high G + C fraction.

In the following section the results of the model are reported.

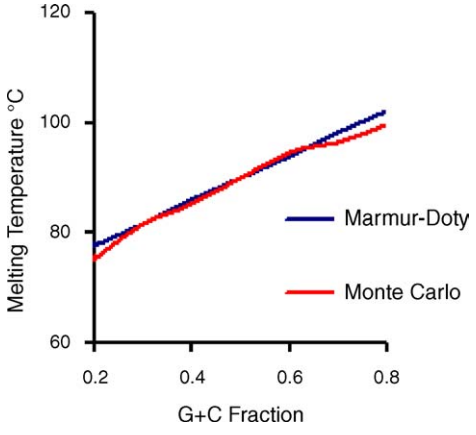

Fig. 2. DNA melting curve.

taccttagcaaatttaaggcgctatcaacgtactggagtggttaacgaaggatcgacccagttgttgtcagaatcctgcagcaatgataataaccatttt
gcttcgcagttattgatagcgtgaatcccgcctatcatgcgtccaccaaggtgtatataaatgtttaatcaaagtctgtgttaaacagagtaactgtaatcc
cacccgtgccttaacctgggtgttggtcagacttcgtaggaacgaccggcaggcggctttatctacaacgtaagcggagtatttcgaaatttgctaatta
ccgatagacgatttatttttcctatagttgcactacctcgttattatacgtttgaagccgctataaagagctgaaaaaagttaccaaattcgtgttttttcggctc
tacatat

Fig. 3. Computer generated sequence of the 3′–5′ strand of target.

3. Results

The theoretical error productions of two PCR protocols are compared in this study. The first protocol (P1) is typical for a conventional heat block thermocycler and the second protocol (P2) is typical for a rapid thermocycler, the PCRJet® (Quintanar and Nelson, 2002). For the purpose of the model a sequence of 416 nucleotides has been generated that serves as the 3′–5′ strand of the template and it is shown in Fig. 3. The G + C content is 40%. The primer lengths are 16 bp each.

The first protocol (P1), proposed by Saiki et al. (1985), has the following holding times per cycle; 1 min at 94 °C, 1 min at 55 °C, 1 min at 72 °C. A PCR cycle consists of the holding times and the ramp times. The temperature–time history of a single cycle is shown in Fig. 4.

The second protocol (P2) is for the fast PCRJet® thermocycler (c.f. review in Nature—Moore, 2005). Fast thermocycling with the PCRJet® is achieved by the flow of high velocity air through a reaction chamber that houses an array of cuvettes. Two inlet air streams (one hot, one cold) are pre-conditioned and mixed to achieve the required temperature and flow rate in the reaction chamber. The response time of the PCRJet® is less than 100 ms, thus users are able to do “thermal management”, which is the concept that DNA damage can be minimized with fast and precise thermal control.

In Fig. 5 a PCRJet® temperature–time profile is shown. The profile is typical for the PCR protocols which are used in PCRJet® experiments to amplify templates of similar length as the one shown in Fig. 3. The maximum 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, a rate of 47 °C/s has been measured. In Table 3 the time intervals for both protocols are listed. Time is measured in milliseconds. The total time for P1 is 206.000 s and for P2 the total time is 4.660 s.

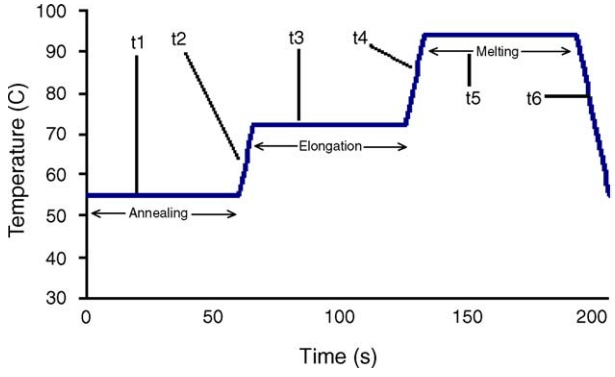

Fig. 4. Thermal cycle of Saiki et al. protocol (P1). The annealing time is $t1$, the ramp time to the elongation temperature is $t2$, the elongation time is $t3$, the ramp time to denature temperature is $t4$, the denature time is $t5$ and the ramp time to the annealing temperature is $t6$.

3.1. Thermal damage for protocol P1

In Fig. 6 the cumulative errors due to A + G thermal depurination are shown for the P1 protocol. The most damage occurs when the DNA is in single-stranded form. If the progression of the error accumulation is overlayed on the temperature–time history of Fig. 4, it is noted that the error rate increases

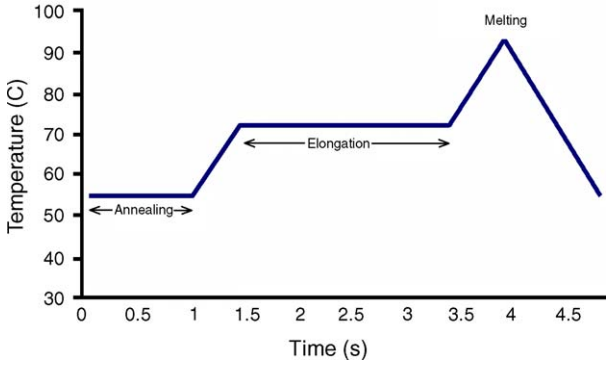

Fig. 5. Temperature (°C) vs. Time (s) cycle for the PCRJet® protocol (P2).

Table 3
Time intervals of protocols P1 and P2 (ms)

	$t1$	$t2$	$t3$	$t4$	$t5$	$t6$	Total
P1	60000	5600	60000	7400	60000	13000	206000
P2	1000	360	2000	470	0	830	4660

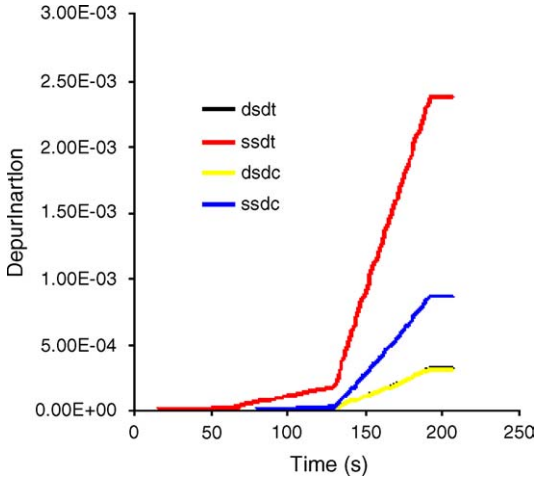

Fig. 6. Cumulative depurination damage over one cycle of protocol P1. The double stranded depurination values for the template and the complementary strand are very close and the plots overlap.

dramatically when the temperature rises from 72 to 94 °C. The extensive denaturation hold largely contributes to the total error count. The linear slopes of the depurination graphs indicate a constant rate of depurination and thus a constant temperature (denaturation value). The complementary strand has fewer depurination errors. The template and its complementary strand have approximately the same number of purine sites (A + G), hence the 'dsdc' and 'dsdt' curves are nearly identical.

In Fig. 7, cytosine deamination in ssDNA and dsDNA is shown over one thermal cycle. Similar to the depurination case, the single-stranded cytosines deaminate the most, followed by the single-stranded complementary strand. If the cytosines are in double-stranded form, the deamination of the two strands is the same since they have equal cytosine content. However, the contribution of the double-stranded deamination to the total deamination is almost negligible. In order to control the C deamination damage, the exposure time of single-stranded DNA to high temperature must be minimized.

A comparison of Figs. 6 and 7 confirms that depurination is the dominant source of thermal DNA damage. In addition, the most damage is incurred when the DNA is in single-stranded form. The total error accumulation in the template after one cycle is 0.003. This number can be interpreted in various ways. When a template of 416 bp is copied, there is a 0.3% chance to have either a depurination (most likely) or C deamination event. On the average, one in every 333 copies will have a thermal error. The thermal error frequency for P1 is expressed as errors/10^6 bp:

$$\Phi_{\mathrm{T}} = 7.2\,\mathrm{errors}/10^6\,\mathrm{bp}$$

The error frequency for the *KOD* polymerase is in the order of 1.1 errors/10^6 nt. The thermal damage contributes far more to the total error count than the polymerase. The model predicts a total error frequency of:

$$\Phi_{\mathrm{Total}} = 8.3\,\mathrm{errors}/10^6\,\mathrm{bp}$$

Thermal damage to DNA is minimized when nucleic acids spend as little time as possible at elevated temperatures (>50 °C). Therefore, thermal damage to DNA is reduced when fast thermocycling protocols are used. In the following section a typical protocol of the PCRJet® is analyzed.

3.2. Thermal damage for protocol P2

In Fig. 5 the temperature–time profile is shown for protocol P2. The sample is annealed for 1 s, the temperature is raised at a rate of 47 °C/s to 72 °C, where it remains for 2 s before it is raised to 94 °C. No time is spent at the denaturation temperature and the sample is immediately cooled to 55 °C. The total time for one cycle is 4.66 s.

In Fig. 8 the cumulative depurination damage is shown for the template DNA and its complementary strand. The rates increase when the temperature is raised after the annealing step. A closer look at the 'ssdt' curve reveals that, towards the end of the elongation step, the rate slows down a little. The explanation is as follows. The template can be divided into two parts: behind the polymerase complex site it is double stranded and partially melted, ahead of the insertion site it is only single stranded. The (A + G) concentration in the part of the template that has not been copied becomes progressively less with extension. The amount of single stranded (A + G) in the part of the template that has been extended depends on the melting behavior of the dsDNA. Over the first 2.5 s the total single stranded (A + G) concentration declines, because the gain in single stranded (A + G) due to melting of dsDNA is more than offset by the reduction in the (A + G) concentration in the single stranded part of the template. The point is illustrated in Fig. 9. The single stranded depurination of the template (denoted as d*D*/d*t*) is plotted as a function of time over one cycle. The rate of depurination is the product of the rate constant (Eq. (8)) and the number of single stranded

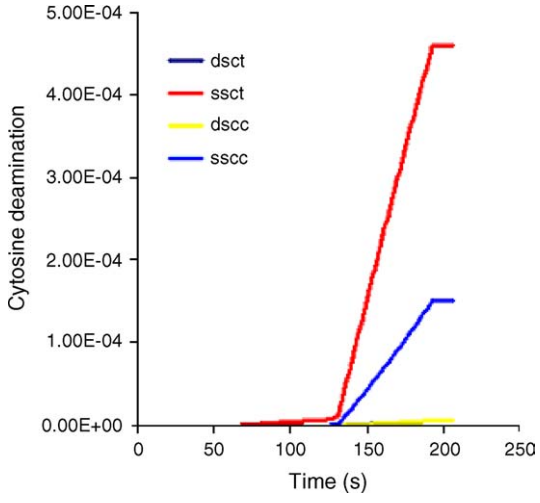

Fig. 7. Cumulative cytosine deamination over one cycle of protocol P1. The double stranded values for the template (dsct) and the complementary strand (dscc) are very close and the plots overlap.

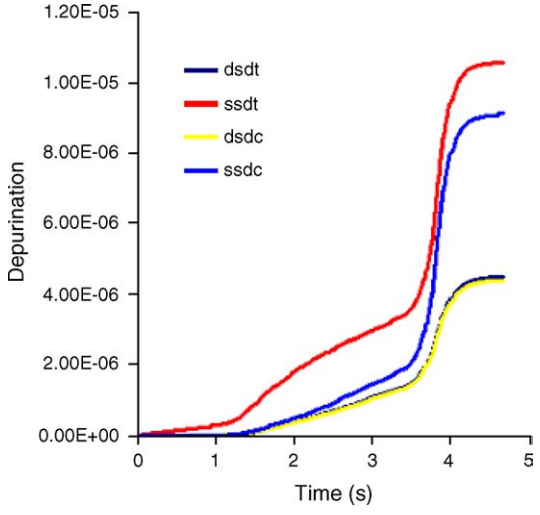

Fig. 8. Cumulative depurination damage over one cycle of protocol P2. The double stranded depurination values for the template and the complementary strand are very close and the plots overlap.

purines:

$$\frac{dD}{dt} = r_{DPss}[A+G]_{ss} = 7.6 \times 10^{11} e^{-31/RT}[A+G]_{ss}$$

The rate in Fig. 9 is scaled with its maximum value at the denaturation temperature. The number of [A + G] sites on the template strand that are single stranded are shown as well, scaled by their maximum. The maximum {A + G} single stranded sites of 202 are at the start of the cycle. The rate has a skew bimodal shape which indicates two local maxima. Over the first 2.5 s the ss [A + G] number decreases, primarily due to a conversion into ds [A + G] as the extension proceeds and not due to depurination—note the low depurination rate over this period. The first maximum of the depurination rate is a result of the relatively high number of ss [A + G] at the time when the temperature increases to the elongation value. The second maximum is due to the maximum (denaturation) temperature. When the temperature increases to the denaturation value, ds [A + G] is converted into ss [A + G]—this increase in ss [A + G] compounds the higher reaction rate constant and the depurination rate shows the sharp spike. If the denaturation time would be prolonged, the rate would decline after ss [A + G] has been depleted.

Fig. 10 shows the C deamination damage for the fast protocol P2. The qualititative forms of the curves are similar to the depurination case, but the values are lower by factor 10. The single stranded C deamination is much larger than double stranded C deamination. It is further noted that the template strand and complementary strand acquire almost the same amount of C deamination at the end of the cycle. The huge increase in C deamination, when the temperature increases to the denaturation value, is a result of the Arrhenius effect.

The total error frequency for P2 is:

$$\Phi_{Total} = \Phi_T + \Phi_{PCR} = (0.026 + 1.1)\,\text{errors}/10^6\,\text{bp}$$

Clearly the PCR fidelity plays the most important role in the error production of protocol P2.

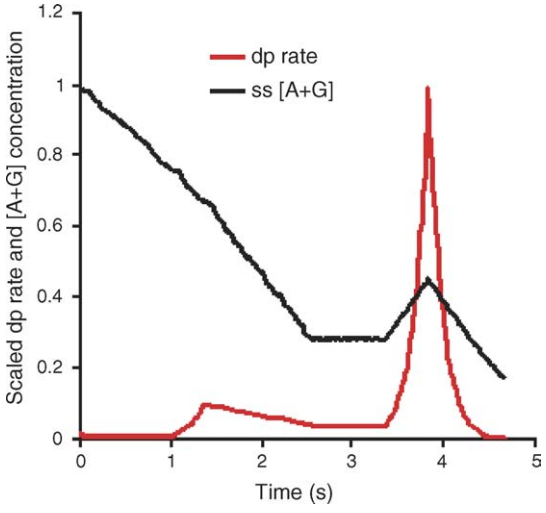

Fig. 9. The single stranded depurination (dp) rate, scaled by the maximum and the number of single stranded [A + G] scaled by its maximum.

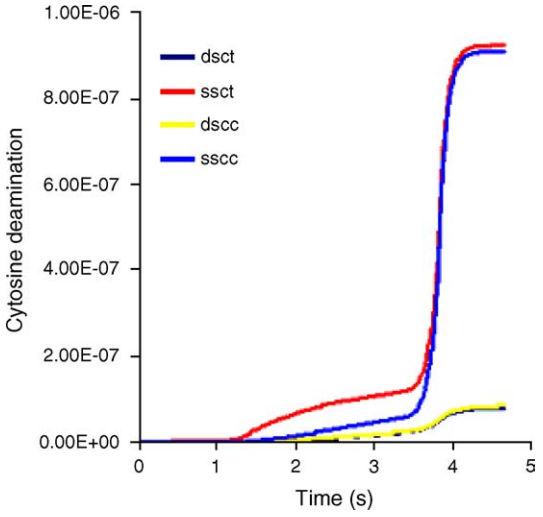

Fig. 10. Cumulative cytosine deamination over one cycle of protocol P2. The double-stranded DNA values (dsct) and the complementary strand (dscc) are very close and the plots overlap.

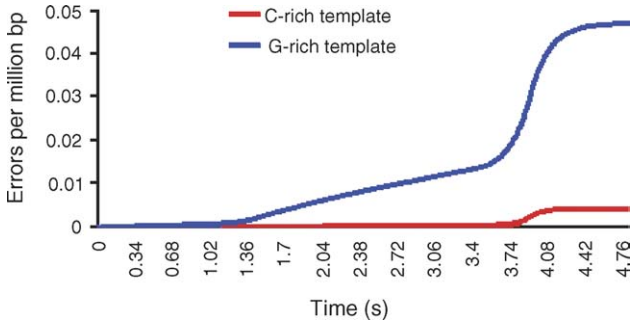

Fig. 11. Cumulative thermal damage for a template that contains 60% G + C. The complementary strand contains 40% C and the template strand contains 40% G.

The target DNA that has been used in all the examples contains 40% (G + C) and this composition holds for the template and complementary strands. An interesting situation arises when the template strand contains 40% G and the complementary strand contains 40% C. The (G + C) content of the target is now 60%, but the purines and pyrimidines are not evenly distributed between the two strands. The total errors (expressed in errors/million base pairs) are shown in Fig. 11.

The template strand acquires significantly more errors than the complementary strand. The reason is that depurination rates are higher than C deamination rates. If the thermal cycle involves long holds at elevated temperatures the disparity in error content between the strands will become more pronounced.

4. Other implications of thermal damage

The effect of thermal damage in DNA reaches beyond classical PCR applications. Nucleotides in the presence of water are subject to hydrolytic attack and higher temperatures accelerate

the reaction rates. We want to mention a few situations where thermal damage is of particular concern.

4.1. Construction of DNA molecules from synthetic oligonucleotides

Although damage does not affect diagnostic PCR too adversely, there are applications of PCR where error minimization becomes important. One example is the construction of large DNA molecules from oligonucleotides. It is crucial that DNA molecules which are assembled from shorter oligonucleotides are error-free or nearly error-free to maintain their functionality. In the PCA strategies originally described by Barnett and Erfle (1990) and Ciccarelli et al. (1991), several overlapping non-phosphorylated 3′OH oligonucleotides are annealed in a single reaction at concentrations of less than 0.1 μM. After 20–30 cycles of PCR amplification, the concentrations of the two outer primers are increased to ∼1 μM and another 30–35 cycles of PCR are carried out (Stemmer et al., 1995). Although only a few full-length DNA molecules assemble during the first PCR stage, they are selectively amplified during the second stage. Using this technique, Stemmer et al. (1995) constructed a ∼2700 bp long synthetic plasmid from 134 different overlapping 3′OH oligonucleotides. Refinements on the Stemmer et al. (1995) protocol have been described by Mehta et al. (1997), Gao et al. (2004), Shevchuk et al. (2004), and Young and Dong (2004), whereby double-stranded DNA molecules of 1.2–20 kb have been assembled by PCA.

In the classic method of gene assembly used by Khorana (1979), synthetic DNA oligonucleotides are 5′ phosphorylated using T4 kinase, annealed to form overlapping duplexes, and then enzymatically joined together using T4 DNA ligase. A variation of this method was recently used by Smith et al. (2003) to synthesize a 5386 bp ΦX174 RFI DNA molecule in 2 weeks. Over 250 oligonucleotides of 42 bp each were 5′ phosphorylated, gel-purified, annealed, and ligated. After overnight high temperature ligation (65 °C), the DNA ligated product was then amplified using the PCA strategy of Stemmer et al. (1995). Therefore, enzymatic steps such as a 65 °C overnight ligase reaction (Smith et al., 2003) are expected to result in substantial thermal depurination and C deamination in DNA; on the order of 2–3 errors/1000 bp.

4.2. PCR amplification of long DNA fragments

Typical amplification conditions for DNA fragments >20 kb would involve an initial denaturing step at 92 °C for 2 min, followed by a two-step PCR regime of 10 cycles at 92 °C for 10 s, 55–65 °C for 30 s, and 68–72 °C for 18 min: 20 cycles at 92 °C for 10 s, 55–65 °C for 30 s and 68–72 °C for 18 min with an additional 10 s/cycle elongation for each successive cycle, concluded by a final elongation at 68–72 °C for 7 min (www.roche-applied-science.com). A three-step cycling is proposed if the T_m of the primers is below 68 °C: pre-amplification denaturation at 94 °C for 30 s, followed by either 30–35 cycles of 94 °C for 30 s, 55–65 °C for 30 s and 68–72 °C for 45–60 s/kb of target. If the T_m of the primers is equal or greater 68 °C, a two-step cycling of 30–35 cycles at 94 °C for 30 s, 68 °C for 45–60 s/kb of target (www.invitrogen.com). If the target molecule has a length of 20 kb, it is recommended to use an annealing or annealing/extension period of 15–20 min/cycle at temperatures between 68 and 72 °C. Amplification of longer templates require higher dNTP concentrations (∼500 μM) and salt concentrations (∼2.75 mM) as well as longer oligonucleotides (22–34 nt) with balancing melting temperatures above 60 °C to enhance reaction specificity (www.roche-applied-science.com).

The typical conditions which have been described here will make the DNA fragments especially vulnerable to thermal damage. The model that has been presented can be used to determine the thermal damage for such long protocols. Also no mention has been made until now of the effect of thermal damage on the primers. These short DNA fragments are subject to the same hydrolytic attack as the DNA templates. Therefore similar modifications of their structure are expected and in time-consuming protocols with prolonged exposure at high temperatures the primer/template annealing will be compromised.

5. Conclusions

The PCR process typically consists of three distinct thermal steps: denaturation at temperatures >90 °C, primer annealing at 50–65 °C, followed by enzymatic elongation at 68–75 °C. At annealing temperatures the DNA molecules are mostly in the double-stranded form, hence the hydrolytic attack on the DNA bases are sterically hindered. Two hydrolytic damage reactions are prominent at elevated temperatures; C deamination and A + G depurination. The rate constants of these thermal damage reactions follow Arrhenius kinetics, hence the rates increase rapidly with temperature. The following conclusions are drawn from this study:

1. Depurination plays a more important role than C deamination in thermal damage to DNA during PCR amplification.
2. If the standard PCR thermocycling protocols (c.f. Saiki et al., 1985) are followed, then thermal damage may be the largest contributor to the overall error frequency in amplified DNA.
3. Thermal damage to DNA is minimized if high-speed PCR protocols are employed.
4. In particular, the denaturation holding time must be kept to a minimum.
5. The periods (time/cycle) for the protocols P1 and P2 are 206 and 4.7 s, respectively. The total errors (depurination and C deamination) are $7.2/10^6$ bp and $0.026/10^6$ bp, respectively. Thus a 50-fold reduction in the PCR period leads to a 270-fold reduction in errors.
6. Thermal damage can help to explain the reduction in yields at high cycle numbers.
7. Long PCR is subject to significant thermal damage. Quantitative models of thermal damage are the ideal tools to investigate the role that thermal damage may play in long PCR and it may lead to a better understanding of the process.
8. New PCR applications, such as assembly PCR, require higher standards of replication fidelity. To attain these higher stan-

dards, current protocols will have to include thermal management.

Acknowledgement

EP and HJV gratefully acknowledge the financial support of the National Institutes of Health through grant R21RR20219.

References

Barnett, R.W., Erfle, H., 1990. Rapid generation of DNA fragments by PCR amplification of crude synthetic oligonucleotides. Nucleic Acids Res. 18, 3094.

Cadet, J., Bellon, S., Berger, M., Bourdat, A.-G., Douki, T., Duarte, V., Frelon, S., Gasparutto, D., Muller, E., Ravanat, J.-L., Sauvaigo, S., 2002. Recent aspects of oxidative DNA damage: guanine lesions measurement and substrate specificity of DNA repair glycosylases. Biol. Chem. 383, 933–943.

Ciccarelli, R.B., Bunyuzlu, P., Huan, J., Scott, C., Oakes, F., 1991. Construction of synthetic genes using PCR after automated DNA synthesis of their entire top and bottom strands. Nucleic Acids Res. 19, 6007–6013.

Cline, J., Braman, J.C., Hogrefe, H.H., 1996. PCR fidelity of *Pfu* DNA polymerase and other thermostable polymerises. Nucleic Acids Res. 24, 3546–3551.

Crothers, D.M., Zimm, B.H., 1964. Theory of the melting transition of synthetic polynucleotides: evaluation of the stacking free energy. J. Mol. Biol. 9, 1–9.

Frank-Kamenetskii, M.D., 1971. Simplification of the empirical relationship between melting temperature of DNA, its GC content and concentration of sodium ions in solution. Biopolymers 10, 2623–2624.

Fryxell, K.J., Zuckerkandl, E., 2000. Cytosine deamination plays a primary role in the evolution of mammalian isochors. Mol. Biol. E 17, 1371–1383.

Gao, X., Yo, P., Keith, A., Ragan, T., Harris, T., 2004. Thermodynamically balanced inside-out TBIO PCR-based gene synthesis: a novel method of primer design for high-fidelity assembly of longer gene sequences. Nucleic Acids Res. 31, e143–e164.

Gelfand, D.H., White, T.J., 1990. Thermostable DNA polymerases. In: Innis, MA., Gelfand, DH., Sninsky, JJ., White, TJ. (Eds.), PCR Protocols: a Guide to Methods and Applications. Academic Press, Inc., San Diego.

Gray, D.M., Tinoco Jr., I., 1970. A new approach to the study of sequence-dependent properties of polynucleotides. Biopolymers 9, 223–244.

Griep, M., Whitney, S.E., Nelson, R.M., Viljoen, H.J., 2006. DNA polymerase chain reaction: prediction of error frequencies and extension rates. AIChE J. 52, 384–392.

Griep, M., Kotera, C.A., Nelson, R.M., Viljoen, H.J., 2005. Kinetics of the DNA polymerase *Pyrococcus Kodakaraensis*. Chem. Eng. Sci., submitted for publication.

Grisham, C.M., Garrett, R.H., 1998. Biochemistry. Thomson Learning.

Gustafson, C.E., Alm, R.A., Trust, T.J., 1993. Effect of heat denaturation of target DNA on the PCR amplification. Gene 123, 241–244.

Hsu, G.W., Ober, M., Carell, T., Beese, L.S., 2004. Error-prone replication of oxidatively damaged DNA by a high-fidelity polymerase. Nature 431, 217–221.

Innis, M.A., Myambo, K.B., Gelfand, D.H., Brow, M.A.D., 1988. DNA sequencing with *Thermus aquaticus* DNA polymerase and direct sequencing of polymerase chain-reaction amplified DNA. Proc. Natl. Acad. Sci. U.S.A. 85, 9436–9440.

Jager, J., Pata, J.D., 1999. Getting a grip: polymerases and their substrates. Curr. Opin. Struct. Biol. 9, 21–28.

Johnson, K.A., 1993. Conformational coupling in DNA polymerase fidelity. Annu. Rev. Biochem. 62, 685–713.

Khorana, H.G., 1979. Total synthesis of a gene. Science 203, 614–625.

Kiefer, J.R., Mao, C., Braman, J.C., Beese, L.S., 1998. Visualizing DNA replication in a catalytically active *Bacillus* DNA polymerase crystal. Nature 391, 304–307.

Li, Y., Waksman, G., 2001. Structural studies of the *KlenTaq*1 DNA polymerase. Curr. Org. Chem. 5, 871–883.

Lindahl, T., Nyberg, B., 1972. Rate of depurination of native DNA. Biochemistry 11, 3610–3618.

Lindahl, T., Nyberg, B., 1974. Heat-induced deamination of cytosine residues in deoxyribonucleic acid. Biochemistry 13, 3405–3410.

Mehta, D.V., DiGate, R.J., Banville, D.L., Guiles, R.D., 1997. Optimized gene synthesis, high level expression, isotopic enrichment, and refolding of human interleukin-5. Protein Expression Purif. 11, 86–94.

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

Moore, P., 2005. PCR: replicating success. Nature 435, 235–238.

Ninio, J., 1987. Alternative to the steady-state method: derivation of reaction rates from first-passage times and pathway probabilities. Proc. Natl. Acad. Sci. U.S.A. 84, 663–667.

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

Quintanar, A., Nelson, R.M., 2002. A process and apparatus for high-speed amplification of DNA. U.S. Patent No. 6,472,186.

Saiki, R.K., Scharf, S., Faloona, F., Mullis, K.B., Horn, G.T., Erlich, H.A., Arnheim, N., 1985. Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science 230, 1350–1354.

Santalucia, J., 1998. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. Proc. Natl. Acad. Sci. U.S.A. 95, 1460–1465.

Shevchuk, N.A., Bryksin, A.V., Nusinovich, Y.A., Cabello, F.C., Sutherland, M., Ladisch, S., 2004. Construction of long DNA molecules using long PCR-based fusion of several fragments simultaneously. Nucleic Acids Res. 32, e19–e47.

Sikorsky, J.A., Primerano, D.A., Fenger, T.W., Denvir, J., 2004. Effect of DNA damage on PCR amplification efficiency with the relative threshold cycle method. Biochem. Biophys. Res. Commun. 323, 823–830.

Smith, H.O., Hutchison, C.A., Pfannkock, C., Venter, J.C., 2003. Generating a synthetic genome by whole genome assembly: ΦX174 bacteriophage from synthetic oligonucleotides. Proc. Natl. Acad. Sci. U.S.A. 100, 15440–15445.

Stemmer, W.P.C., Crameri, A., Ha, K.D., Brennan, T.M., Heyneker, H.L., 1995. Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. Gene 164, 49–53.

Uhlenbeck, O.C., Borer, P.N., Dengler, B., Tinoco Jr., I., 1973. Stability of RNA hairpin loops: a 6-C m-U 6. J. Mol. Biol. 73, 483–496.

Viljoen, S.V., Griep, M., Nelson, R.M., Viljoen, H.J., 2005. A macroscopic kinetic model for DNA polymerase elongation and high fidelity nucleotide selection. Comput. Biol. Chem. 29, 101–110.

Whitney, S.E., Sudhir, A., Nelson, R.M., Viljoen, H.J., 2004. Principles of rapid polymerase chain reactions: mathematical modeling and experimental verification. Comput. Biol. Chem. 28, 195–209.

Yap, E.P., Mcgee, J.O., 1991. Short PCR product yields improved by lower denaturation temperatures. NAR 19, 1713.

Young, L., Dong, Q., 2004. Two-step total gene synthesis method. Nucleic Acids Res. 32, e59–e71.

Further reading

Web references

www.invitrogen.com
www.roche-applied-science.com.