Replication of the Base Pair 6-Thioguanine/5-Methyl-2-pyrimidinone with the Large Klenow Fragment of *Escherichia coli* DNA Polymerase I

Harry P. Rappaport

Biology Department, Temple University, Philadelphia, Pennsylvania 19122

Received June 16, 1992; Revised Manuscript Received December 8, 1992

ABSTRACT: The kinetics and the fidelity of replication of the base pair 6-thioguanine (G^s)/5-methyl-2-pyrimidinone (T^h) have been determined by using defined oligomers with the large Klenow fragment of *Escherichia coli* DNA polymerase I. The insertion efficiency, V_{max}/K_m (min^{-1} μM^{-1}), of T^h opposite G^s is 1.5, and the insertion efficiency of G^s opposite T^h is 0.7. By comparison, the insertion efficiencies of C opposite G and G opposite C are 0.5 and 1.5. The insertion efficiency of the next base, A opposite T, is 2 times greater after the base pair G^s/T^h than after G/C. The fidelity of replication with respect to thymine and adenine has misinsertion frequencies, or ratios of the insertion efficiency of the "wrong" base to the "right" base, of 7×10^{-4} for T opposite G^s (T/G^s), 4×10^{-6} for T/T^h, and a maximum stable misinsertion frequency of 4×10^{-4} for A/T^h. No detectable elongation occurs after an A is inserted opposite a G^s. These values are similar to the misinsertion frequencies of G and C with T and A. The maximum stable misinsertion frequencies with G and C are 4×10^{-2} for G/T^h, 3×10^{-2}–7×10^{-3} for G^s/C, and 2.6×10^{-1} for C/G^s, and the misinsertion frequency is $<1 \times 10^{-3}$ for T^h/G. The kinetics results and molecular modeling suggest modifications to the G^s/T^h base pair that may provide higher levels of fidelity of replication with respect to C and G.

Three reports of investigations of alternative base pairs to the standard base pairs have appeared recently in the literature (Rappaport, 1988; Switzer et al., 1989; Piccirilli et al., 1990). This is the second report of investigations of a class of alternative base pairs that have sulfur and hydrogen as complementary units in place of oxygen and the amino group. The previous paper (Rappaport, 1988) provided estimates of the relative association constants of oligomers containing the bases 6-thioguanine (G^s)[1] and 5-methyl-2-pyrimidinone (T^h) with each other and with the standard bases. The interactions of 6-thioguanine and 5-methyl-2-pyrimidinone were very similar to the interactions of adenine and thymine. Figure 1 shows a schematic representation of the structure of the two bases as they might appear in a base pair. The present investigation addresses the questions of the kinetics and fidelity of replication using the large Klenow fragment of *Escherichia coli* polymerase I.

Switzer et al. (1989) and Piccinilli (1990) have published accounts of their investigations of two alternative base pairs using oxygen and the amino group in different arrangements than in the standard bases. Since no detailed kinetics parameters were published, it is not possible to evaluate the kinetics reasons for the fidelity levels obtained with the Klenow fragment.

A detailed kinetics model resulting from several investigations of the replication and fidelity mechanisms of the Klenow fragment has been published (Kuchta et al., 1988). Boosalis et al. (1987) have provided a theoretical procedure for evaluating the kinetics parameters of polymerization when no exonuclease activity is present and several deoxynucleotides are added. Taken together, the papers suggest that it should be possible to arrange the assay conditions for polymerase and exonuclease activities so that the velocity of each reaction is proportional to the concentration of the template–primer used in the reaction. This would make the mathematical discription of the kinetics equivalent to a Markov chain in probability theory (Van Kampen, 1981). This expectation was realized in the present investigation.

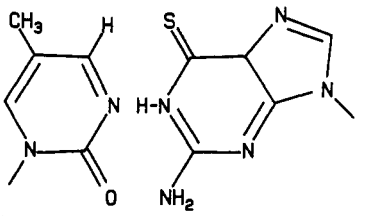


FIGURE 1: Schematic representation of the base pair 5-methyl-2-pyrimidinone, left base, and 6-thioguanine, right base.

MATERIALS AND METHODS

Materials

Enzymes were obtained from United States Biochemicals (T4 polynucleotide kinase), New England Biolabs (T4 ligase, large Klenow fragment of *Escherichia coli* polymerase I), and Sigma (bacterial alkaline phosphatase, snake venom phosphodiesterase). Chemicals were obtained from Sigma (all standard deoxynucleotide triphosphates), Aldrich Chemical Co. (all anhydrous solvents and reagents for synthesis), and ICN Radiochemicals (adenosine 5′-triphosphate (γ-^{32}P)). HPLC columns were obtained from Synchropak, Lafayette, IN 47902. TLC and PLC plates were obtained from Whatman and Merck.

Methods

HPLC. All studies used a Milton Roy fixed wavelength monitor D at either 254 or 326 nm. The noise level with flow was less than 10^{-4} AU.

Electrophoresis. Preparative and analytical electrophoresis used 20–25% acrylamide with 7 M urea in buffer containing

[1] Abbreviations: G^s, 6-thioguanine; T^h, 5-methyl-2-pyrimidinone; dNTP, deoxynucleoside 5′-triphosphate; KF, large Klenow fragment of *Escherichia coli* DNA polymerase I; EDTA, ethylenediaminetetraacetate, sodium salt; pp_i, inorganic pyrophosphate; Tris-HCl, tris(hydroxymethyl)-aminomethane hydrochloride; HPLC, high-pressure liquid chromatography.

0006-2960/93/0432-3047$04.00/0

Table I: Primers and Templates Synthesized

designation	sequence
template	
T-1	5′-T-T-T^{h}-A-C-G-G-A-C-C-C-G-T-C
T-2	5′-A-A-A-G-T-A-G-C-G-C-T-A
T-3	5′-(P-1)-T-G^{s}-(T-2)
T-4	5′-T-C-G-A-C-G-G-A-C-C-C-G-T-C
T-5	5′-G-T-C-A-C-G-G-A-C-C-C-G-T-C
primer	
P-1	5′-G-A-C-G-G-G-T-C-C-G
P-2	5′-T-A-G-C-G-C-T-A-C-T

50 mM Tris–borate and 10^{-3} M EDTA, pH 8.3. When oligomers containing 6-thioguanine were purified by electrophoresis, 1 mM glutathione was included in the cathode buffer, and presample electrophoresis was run for 30 min.

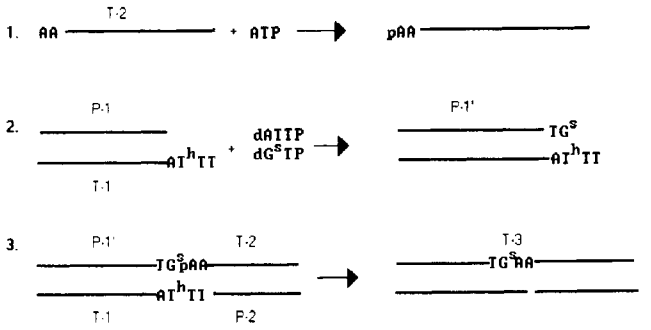
Autoradiography and Gel Counting. After electrophoresis of radioactive oligomers, the gel was covered with Saran Wrap and overlaid with Kodak X-ray film (GPB-1). The film was exposed at −20 °C under uniformly distributed weight for 12–48 h. After removal of the film, the gel was allowed to thaw under weights. Radioactive india ink marks on the Saran Wrap were used for alignment, and a paper tracing of the developed X-ray film was made. The tracing was attached to the Saran Wrap covering the gel, using the india ink lines on the Saran Wrap for alignment. The radioactive bands were excised and were counted using Cerenkov radiation.

Oligodeoxynucleotide Synthesis and Purification. The synthesis of the deoxynucleosides of 6-thioguanine and 5-methyl-2-pyrimidinone followed published methods (Roark et al., 1978; Rappaport, 1988). With the exception of the oligomer containing 6-thioguanine, all template and primer oligomers were synthesized by the phosphite triester method (Matteucci & Caruthers, 1981). The compositions of the oligomers are listed in Table I.

5-Methyl-2-pyrimidinone Oligomer. The preparation of the 3′-*O*-[(*N*,*N*-diisopropylamino)(2-cyanoethoxy)phosphinyl]-5′-*O*-(4,4′-dimethoxytrityl)-2′-deoxyribosyl-5-methyl-2-pyrimidinone essentially followed the method of McBride and Caruthers (1983). The sequence of the oligomer was TTT^{h}-ACGGACCCGTC. The first 11 nucleotides were added with an automatic synthesizer. The last four nucleotides were added by using the syringe method of Tanaka and Letsinger (1982). After complete deprotection the major oligomer was isolated either by acrylamide electrophoresis or by PLC using Whatman PK 6F plates, 1000 μm (Michniewicz et al., 1973). The oligomer was purified on a C_{18} reversed-phase column, 250 × 4.6 mm, using 0.02 M KH_2PO_4, pH 5.5, and a gradient of methanol (Gilder & McLaughlin, 1989). In agreement with Gilder and McLaughlin, two significant peaks were found. The second peak had the appropriate adsorption spectra between 310 and 340 nm for the presence of 5-methyl-2-pyrimidinone. Digestion with snake venom phosphodiesterase and bacterial alkaline phosphatase and HPLC analysis confirmed the presence of the correct amount of deoxyribosyl-5-methyl-2-pyrimidinone.

6-Thioguanine Oligomer. Because the determination of the amounts of deoxynucleosides in an oligomer by HPLC analysis is uncertain to about 10%, it is not possible to detect a 5–10% conversion of 6-thioguanine to guanine during the synthesis and deprotection of an oligomer with one thioguanine base and several guanine bases. A 5% conversion would produce an apparent significant amount of misinsertion of C opposite G^{s}. Because of this concern, an enzymic procedure was adopted for the synthesis of the oligomer containing G^{s}. Since no chemical workup of the oligomer takes place and

Scheme I



since no dGTP is present, no substitution of guanine for thioguanine should take place. The sole source of guanine contamination would be from the source of 6-thioguanine. HPLC analysis of the deoxy-6-thioguanosine source did not detect deoxyguanosine, which placed a bound on deoxyguanosine of less than 0.01%.

The synthesis had three enzymic steps, shown in Scheme I. The first step was the phosphorylation (^{32}P) of the 5′ end of oligomer T-2 using T4 polynucleotide kinase. The second step was to extend the standard primer P-1 by the addition of T and G^{s}, using the template T-1 and KF. The third step ligated the extended primer, still annealed to the T-1 template, to T-2 annealed to the P-2 primer using T4 DNA ligase. The final product, lightly labeled with [^{32}P]phosphate as the link between T-2 and the extended primer, was purified by electrophoresis.

Control experiments showed no detectable synthesis of any polymer with a length near the product length if deoxy-6-thioguanosine 5′-triphosphate was omitted from step 2. Detailed experiments (see Results) showed the misinsertion frequency of T compared to G^{s} opposite T^{h} to be 4×10^{-6}. There was a very small amount of ligation of T-2 to the T-1 template when unextended P-1 was used.

The detailed protocols are given below. All reactions were carried out in siliconized Eppendorf tubes. The numbering of the protocols corresponds to the enzymatic steps shown in Scheme I.

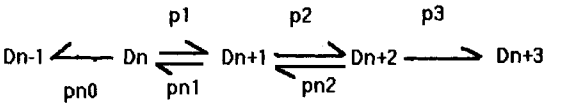
(1) The reaction solution had a volume of 50 μL and contained 50 pmol of T-2 oligomer, 1.8 μM labeled ATP (225 Ci/mm), 5 units of T4 polynucleotide kinase, 60 mM Tris-HCl, pH 7.6, 5 mM dithiothreitol, and 10 mM $MgCl_2$. After incubation at 35 °C for 30 min, an additional 5 units of enzyme and 5 μL of 100 μM ATP (cold) were added, and the incubation was continued for 30 min at 35 °C. The extent of the reaction was followed by TLC with silica gel, using cold ATP as the carrier and water as the solvent. The oligomer remains at the origin, and the ATP moves near the solvent front. The labeled T-2 oligomer was purified on a Waters C_{18} Sep-Pak column. (2) A 30-μL solution containing 50 pmol of the T-1 template, 50 pmol of P-1, 0.09 M Tris-HCl, pH 7.6, 8.33 mM dithiothreitol, and 16.6 mM $MgCl_2$ was heated to 65 °C for 2 min, cooled to room temperature over 10 min, placed on ice, and centrifuged to return all the water to the solution. A 20-μL solution containing 25 μM thymidine 5′-triphosphate, 25 μM deoxy-6-thioguanosine 5′-triphosphate, 125 μg of bovine serum albumin, and 2.5×10^{-2} unit of KF was added to the 30-μL solution. The combined solution was incubated at 20 °C for 30 min, heated to 65–70 °C for 5 min, and cooled to room temperature over 10 min. (3) A 40-μL solution containing the phosphorylated T-2 oligomer, 50 pmol of P-2 oligomer, 60 mM Tris-HCl, pH 7.5, 12 mM $MgCl_2$, and 12 mM dithiothreitol was heated to 65 °C for 2 min, cooled to

room temperature over 10 min, placed on ice, and centrifuged to return all the water to the solution. A 10-μL solution containing 5 mM ATP and 12 weiss units of T4 DNA ligase was added to the 40-μL solution, and the combined 50-μL solution was added to the 50-μL solution from protocol 2. The solution was incubated at 20 °C for 24 h. Two microliters of the solution was added to 8 μL of 90% formamide containing 1 mM EDTA and 0.06% xylene cyanol dye and analyzed by electrophoresis and radioautography. If a yield of greater than 60% was obtained, the rest of the solution was electrophoresed. The electrophoresis was terminated after the dye had moved 9 cm. After exposing X-ray film to the gel at −20 °C, the band with the oligomer was excised from the gel. The excised gel was crushed and extracted with a total of 1 mL of 0.05 M sodium acetate containing 1 mM EDTA and 10 mM mercaptoethanol. After evaporation of most of the water, 250 μL of water and 150 μL of 1 M sodium acetate containing 1 mM EDTA was added. One milliliter of 95% ethanol was added for every 100 μL of solution, and the solution was placed at −20 °C overnight. The solution was centrifuged for 15 min in an Eppendorf centrifuge, the liquid removed, and the tube washed several times with cold 95% ethanol. The precipitated oligomer was dissolved in distilled water with 0.3 mM mercaptoethanol and stored at −20 °C.

Synthesis and Purification of Deoxynucleoside 5′-Triphosphates. 2′-Deoxy-6-thioguanosine 5′-Triphosphate. The procedure of Maybaum et al. (1987) was followed for the synthesis. Dimethylformamide was the solvent for the tributylammonium pyrophosphate. The progress of the reaction was followed by TLC with PEI-cellulose and 1 M LiCl as the solvent. After evaporation of the reaction solution, the residue was taken up in water and desalted with a C_{18} Waters Sep-Pak column. The triphosphate was purified by HPLC using a wax column, 250 × 4.6 mm. The elution was started with 0.1 M ammonium bicarbonate, pH 8.0, 1 mL/min for 7 min. One major peak with absorption at 254 nm was obtained. A linear gradient was started from 0.1 to 1.5 M ammonium bicarbonate. Three or four peaks were obtained. The major peak was the second peak that eluted at 12–13 minutes. The very unique pH-dependent spectrum of 6-thioguanine between 258 and 360 nm was used to verify the presence of 6-thioguanine (Fox et al., 1958). An HPLC analysis of the time course of a bacterial alkaline phosphatase digestion of the peaks established that the first peak was the monophosphate and the second was the triphosphate. The ammonium salt of the triphosphate was changed to the triethylammonium salt. The amount of the triphosphate was determined, using an extinction coefficient of 2.5×10^4 at 340 nm and pH 5. Portions were lyophilized and stored at −20 °C.

2′-Deoxyribosyl-5-methyl-2-pyrimidinonyl 5′-Triphosphate. The 5′ monophosphorylation followed the procedure of Yoshikawa et al. (1967). The monophosphate was purified by HPLC on a wax column, 250 × 4.6 mm, with a linear gradient of triethylammonium bicarbonate, pH 7.6, from 0.02 to 0.05 M. The triphosphate was synthesized from the monophosphate by the method of Hoard and Ott (1965). The proportions of reagents and the method of analysis followed those of Freist and Cramer (1978). The triphosphate was purified by HPLC using a wax column, 250 × 4.6 mm. The elution started with 0.05 M ammonium bicarbonate, pH 7.5, for 12–14 mL, 1 mL/min. A linear gradient was started from 0.05 to 0.5 M ammonium bicarbonate. Three peaks were obtained, with the second peak having most of the 314-nm absorption. The very characteristic spectrum of 5-methyl-2-pyrimidinone in the 258–340-nm range was used to verify the presence of 5-methyl-2-pyrimidinone (Laland & Serck-Hansen, 1964). HPLC analysis of the time course of a bacterial alkaline phosphatase digestion of the second peak confirmed that it was the triphosphate. The triphosphate was converted to the triethylammonium salt, dried, and stored at −20 °C. The amount of the triphosphate was determined using an extinction coefficient of 6×10^3 at 314 nm and pH 6.

Scheme II



General DNA Polymerization Reaction. All reactions took place in siliconized 1.5-mL polypropylene tubes with caps. All solution transfers were performed with Gilson Microman devices. These devices do not use an air cushion. Mixing was done by repeated pipetting of a partial volume of the solution into the remaining solution. Typical final volumes of the reaction mixtures were 10–20 μL. The polymerization reaction was started by mixing two solutions of equal volume. One solution contained the template annealed to the [5′-^{32}P]-phosphate-labeled primer and the polymerase. The second solution contained the deoxynucleoside triphosphates. Typical final concentrations were 60 mM Tris-HCl, pH 7.6, 5 mM dithiothreitol, 10 mM $MgCl_2$, 12 nM template, 8 nM primer, 0.5×10^{-3} unit of KF/10 μL, 25 μg of bovine serum albumin, 2 μM thymidine 5′-triphosphate, and an appropriate concentration of a deoxynucleoside triphosphate. A protocol for annealing the primer to the template: Combine 10 μL each of 0.6 M Tris-HCl, pH 7.6; 50 mM dithiothreitol; 100 mM $MgCl_2$; template (0.12 pmol/μL); and primer (0.08 pmol/μL). Heat the solution to 65 °C for 2 min, and cool the solution to room temperature over 10 min. Place the solution on ice and centrifuge to return all the water to the solution. The temperature for the reaction was 18.5 °C. One-microliter portions were removed at various times and added to 8 μL of 90% formamide containing 1 mM EDTA and 0.06% xylene cyanol dye for electrophoretic analysis.

RESULTS

Introduction For the determination of the kinetics parameters and the degree of fidelity of replication of the base pair G^s/T^h it was desirable to have an assay that would allow the evaluation of the contribution of a 3′–5′-exonuclease activity. The investigation by Kuchta et al. (1988) of KF suggested that assay conditions could be found that would allow the velocity of each reaction to be proportional to the concentration of template–primer used in the reaction, and thus the determination of kinetics parameters for both polymerase and exonuclease activities could be accomplished with one reaction assay if the time course of the production of product was followed. The methodology introduced by Boosalis et al. (1987), which uses the formation of a standard base pair for the first extension step and the formation of the base pair of interest as the second extension step with the quantification of primer extension by gel electrophoresis, was adopted.

The general scheme of the enzymic reactions is shown in Scheme II. The D_n's are the various template–primer complexes with primer length n, the p's are polymerase rates, and the pn's are exonuclease rates. A temperature was sought that would make pn0 negligible compared to p1. In principle

this is not necessary, but it greatly reduces the amount of effort in quantifying the various products. Preliminary experiments indicated that a temperature of about 19 °C would be appropriate. In addition, the T/A base pair obtained at D_{n+1} had a pn1 of less than 10% of p1 at the concentration of dTTP used and could be neglected in the analysis.

Mathematical Descriptions and Experimental Conditions. In order to fulfill the objective of having the velocity of extension of each template–primer species be proportional to the concentration of the template–primer over some range of concentrations, exploratory measurements were made of the addition of a single nucleotide, dTTP, to the template–primer T-4/P-2, Table I. The exploratory measurements showed that if the template–primer concentration was 10 nM or less, then the velocity of addition of the nucleotide to the primer was proportional to the concentration of the template–primer. This was demonstrated experimentally by the exponential loss in time of the initial template–primer and the exponential rise in time of the lengthened primer. That is, the fraction of the initial template–primer, $f_n(t)$, as a function of time was

$$f_n(t) = \exp[(-\mathrm{p1})t]$$

where p1 is a constant in time. In addition, the dependence of p1 on the concentration of dTTP was

$$\mathrm{p1} = \frac{V_{1\mathrm{map}}}{1 + K_{1\mathrm{map}}/[\mathrm{dTTP}]}$$

These equations held for a range of $f_n(t)$ from 1 to 0.3 within 15%. A set of measurements was made of the amount of extended template–primer as a function of time between 1 and 5 min. The amount of the extended primer increased linearly with time. An analysis of this initial velocity as a function of dTTP concentration gave $V_{1\mathrm{map}}$ and $K_{1\mathrm{map}}$ values within 15% of the values obtained from an analysis of p1. If no dTTP was added, then the slow loss of the initial template–primer due to exonuclease activity could be followed. This loss had the form

$$f_n(t) = \exp[(-\mathrm{pn0})t]$$

where pn0 is a constant in time.

The next step was to add a second nucleotide after dTTP and to determine the time dependence of the production of the template–primer D_{n+2}, Scheme II. With the template–primer T-4/P-1, the next standard nucleotide was dGTP. If the concentration of dGTP was chosen so that the rate of its addition was rate limiting, 0.1 μM or less, then the experimental result for the fraction of template–primer with G added, $f_{n+2}(t)$, as a function of time was

$$f_{n+2}(t) = 1 - \exp[(-\mathrm{p2})t]$$

where p2 is a constant in time. The experimental dependence of p2 on the dGTP concentration was

$$\mathrm{p2} = \frac{V_{2\mathrm{map}}}{1 + K_{2\mathrm{map}}/[\mathrm{dGTP}]}$$

The experimental results for two sets of measurements are shown in Figure 2A,B. The low values of p2 for dGTP concentrations of 0.1 and 0.05 μM in one experiment were not reproducible. Experimental points for 0.025 μM from the second experiment (solid symbols) have not been placed on the graph; they overlap the points from the first experiment. The range of $f_{n+2}(t)$ was from 0.05 to 0.7. To determine the relationship of the values of $V_{2\mathrm{map}}$ and $K_{2\mathrm{map}}$ obtained from p2 to values obtained from an initial velocity determination,

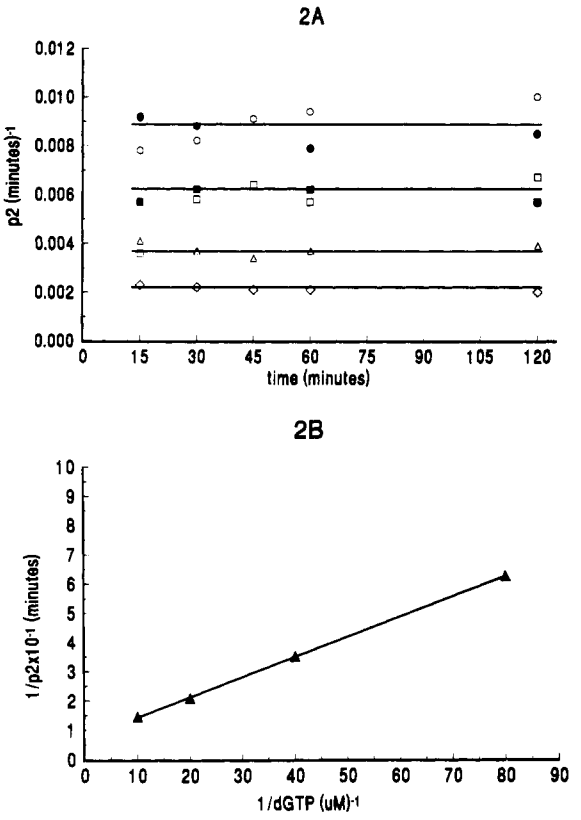


FIGURE 2: Analysis of the concentration dependence of the rate of insertion of G opposite C with KF. (A) The rate constant p2, calculated from eq 1 with $r = 0$, plotted against t for dGTP concentrations of 0.1 (○), 0.05 (□), 0.025 (△), and 0.0125 μM (◇). The solid symbols are from a second set of measurements. The 0.025 μM values from the second set of measurements are not shown because of overlap with the first set. The lines represent the average value of p2 at each concentration. (B) A plot of 1/p2 vs 1/[dGTP] from the data in panel A. Each point is the average of the five time points at each dGTP concentration.

a template–primer with T already added was used. Measurements were made of the production of D_{n+2} from 1 to 5 min, where the amount of D_{n+2} is linear in time. The $V_{2\mathrm{map}}$ and $K_{2\mathrm{map}}$ values determined from the initial velocity were within 12% of the $V_{2\mathrm{map}}$ and $K_{2\mathrm{map}}$ values determined from p2.

As a final control, the concentration of the template–primer was lowered by a factor of 5. This was about the limit that the specific activity of the primer allowed for an acceptable number of counts. The $V_{2\mathrm{map}}$ and $K_{2\mathrm{map}}$ values found from a p2 analysis were within 20% of the previous values. This is an important control because eqs A11, Appendix A, show that the validity of p2 being a constant in time depends on most of the polymerase being in the free form, that is, only a small fraction is present as intermediate. A change of the template–primer concentration by a factor of 5 certainly affects the amounts of intermediates. Since the concentration change did not affect significantly p2, this shows that the free enzyme is the dominant term in eqs A11, Appendix A.

All of these experimental results can be described as the results of a mathematical model where the velocity of each extension of template–primer is proportional to the concentration of template–primer used and that the exonuclease velocity is proportional to the concentration of template–primer degraded. The derivation is given in Appendix A. Three kinetics models of the enzyme-catalyzed polymerization have the property that the rate of each extension of template–primer is proportional to the template–primer concentration. One kinetics model has a sufficient excess of polymerase so

that every template–primer molecule has an enzyme molecule bound. The present investigation does not use this condition. The other two kinetics models depends upon two experimentally controlled conditions: (1) The concentration of the total template–primer is sufficiently low so that each steady-state solution for the velocity of extension of the template–primer is proportional to the concentration of template–primer. (2) The total template–primer concentration is sufficiently low so that the sum of all the concentrations of the enzyme–substrate intermediates is much smaller than the total enzyme concentration. Kinetics model I assumes that the dominant factor determining the rate of formation of the complex between polymerase and template–primer D_{n+1} is the rate of combination of D_{n+1} with the polymerase and not the rate of formation of the complex from the extension step without dissociation. That is, because of the low concentration of the second deoxynucleoside triphosphate, the processive characteristic of the polymerization is lost. Kinetics model II assumes that because the rate of addition of the second nucleotide is rate limiting, due to its low concentration, an equilibrium exits between the polymerase and D_{n+1}. Each of the kinetics models has a different dependence of the apparent V_{max} and K_m on the rate constants. Both kinetics models have the same dependence of V_{max}/K_m on the rate constants. The detailed derivations are given in Appendix A.

In the model, to reproduce the experimental results, it is necessary for the second extension to be rate limiting, just as it was necessary experimentally to obtain a simple exponential buildup of D_{n+2} in time. The equation resulting from the model for p2 is

$$p2 = -\ln\{1 - (r + 1)[f_{n+2}(t)]\}/(r + 1)t \quad (1)$$

where $f_{n+2}(t)$ is the fraction of template–primer with a primer of $n + 2$ nucleotides, p2 is the rate (min^{-1}), t is the time, and r is the ratio of the rate constants, pn2/p2. In Appendix A, eq A7 of kinetics model I and eq A9 of kinetics model II identify p2 as

$$p2 = \frac{V_{2map}}{1 + K_{2map}/[N_2TP]} \quad (2)$$

where V_{2map} is the maximum velocity with a dimension t^{-1}, K_{2map} is the Michaelis–Menten constant, and $[N_2TP]$ is the deoxynucleoside triphosphate concentration for the addition of the second nucleotide. These are apparent values whose explicit definitions in terms of rate constants are given in eq A7, Appendix A. The important parameter V_{2map}/K_{2map} is given in terms of the "true" parameter by eqs A8 and A9, Appendix A,

$$V_{2map}/K_{2map} = V_{2m}/(K_{2D}K_{2m})$$

where K_{2D} is the dissociation constant of D_{n+1} from the enzyme.

In Appendix A, eq A6 gives the rate of formation of D_{n+2} that would be measured if an initial velocity measurement was made. It is clear from the equation that there are two separate terms that give the rate of D_{n+2} production. The experimental result that the initial velocity measurement gave the same values of V_{2map} and K_{2map} as the velocity at later times implies that the dominant term is the $1/[D_{n+1}]$ term in the denominator. Equations A11 of Appendix A require that the free enzyme be the dominant form if p2 is to be a constant in time. The experimental result that a 5-fold decrease in the concentration of template–primer did not change the values of V_{2map} and K_{2map} obtained from an analysis of p2 confirms that the concentration of template–primer was low enough so that the concentrations of enzyme–substrates can be neglected compared to the concentration of free enzyme, and the free enzyme concentration is essentially the total enzyme concentration.

Determination of V_{2map}, K_{2map}, and pn2. For each base pair investigated, an initial survey was performed using a range of concentrations of the deoxynucleoside triphosphate to be added to the primer, step p2, Scheme II, after the addition of the standard deoxynucleoside triphosphate, dTTP, step p1, Scheme II. The range was about 5–0.01 μM. In some cases the range started as high as 40 μM. Samples were taken from the reaction solutions at each concentration of deoxynucleoside triphosphate over a time range of 15–120 min. After electrophoresis to separate the different lengths of primer and extended primer, the bands of radioactivity were excised and counted (details of the excision and counting are given in Materials and Methods under autoradiography and Gel Counting). The total number of counts in the three bands were used to assign the fraction of counts in each band and, thus, the fraction of the initial primer in each band. For a particular reaction with a fixed concentration of the deoxynucleoside triphosphate, the right side of eq 1 with $r = 0$ was plotted against t. The object was to find the approximate highest concentration of dNTP for which the right side of eq 1 is constant or decreases slowly in time. This determines the highest concentration that satisfies the condition that the rate of addition of the deoxynucleoside triphosphate is rate limiting. In addition, the initial survey gave a crude estimate of the apparent K_{2map}. With this information the detailed determinations used an 8–10-fold range of concentrations of the deoxynucleoside triphosphate that had its highest concentration around K_{2map}. The times and the enzyme concentrations were chosen, when possible, so that 40% or more of the primer was extended to the final product by the final time. In each experiment the relative amount of total enzyme used in the reactions was determined by running a reaction with only the standard deoxynucleoside triphosphate, dTTP, present at a fixed concentration. Equations A4 and A5, Appendix A, show that the experimentally observed exponential loss in time of the initial primer has a p1 value proportional to the total enzyme concentration. The dependence of p1 on total enzyme concentration was also verified experimentally. The value of p1 was determined by plotting $-[\ln f_n(t)]/t$ against t, eq A1a, Appendix A. A standard value of p1 was assigned. All values of V_{2map} for the deoxynucleoside triphosphates and the exonuclease velocity, pn2, were scaled to the standard by multiplying by p1(standard)/p1.

To determine the values of p2 and r, the procedure was to calculate the right side of eq 1 with $r = 0$ for all the time points of a particular concentration of the deoxynucleoside triphosphate. If too high a concentration of template–primer was used, then the calculated values of p2 from eq 1 would start low and increase slowly over time as $1/[D_{n+1}]$ became the dominant factor in eq A6, Appendix A. If the expression remained constant for all the time points, then the value of r was below the level of detection, about $0.2V_{2map}$, and the value of the expression was p2. If the expression decreased in time, then the effect of the exonuclease on D_{n+2} was being detected. The function $h(r)$ (eq 3) was plotted against r for

$$h(r) = \frac{\ln\{1 + (r + 1)\,[f_{n+2}(t_1)]\}}{t_1} - \frac{\ln\{1 + (r + 1)\,[f_{n+2}(t_2)]\}}{t_2} \quad (3)$$

all pairs of time points to determine the value of r that makes the expression 0. The values of r for different pairs were

Table II: Results of Calculation of r^a and p2[b] for the Insertion of G opposite T^h When [dGTP] = 1 μM

time pair,[c] min	r	p2	time pair,[c] min	r	p2
15/30	0.51	0.017	30/60	0.8	0.019
15/45	0.97	0.017	30/120	0.8	0.019
15/60	0.75	0.018	45/60	0.4	0.014
15/120	0.81	0.018	45/120	0.8	0.017
30/45	1.2	0.023	60/120	0.8	0.02

[a] Ratio of the rates pn2/p2 in Scheme II and the Results section. [b] Rate of conversion of D_{n+1} to D_{n+2} in Scheme II and the Results section. [c] Assay time t_1/assay time t_2.

averaged. For each set of values of r, f_{n+2}, and t, a p2 value was calculated from eq 1. The p2 values were averaged. The value of pn2 was calculated as r(p2). Equation A9, Appendix A, shows that pn2 is equal to V_{max}/K_m for the exonuclease. The calculations were repeated for each concentration of deoxynucleoside triphosphate. The values of 1/p2 were plotted against $1/[N_2TP]$, and the method of Cleland (1967) was used to determine a straight line and the parameters V_{2map} and K_{2map}. An example of the procedure is shown in Figure 2A,B. Table II shows an example of the calculations of r and p2 for the insertion of G opposite T^h, with primer P-1 and template T-1. The value of pn2 is 0.014. This value has not been scaled to the standard p1. The results of these measurements are presented in Table III.

V_{3map} and K_{3map} from an Extension to D_{n+3} by dATP. In order to determine the effect of the new base-pair combinations on replication, the kinetics parameters for the formation of a standard base pair after the new combinations were determined. The standard deoxynucleoside triphosphate was dATP for the base pair A/T. To be able to use the same methodology for this investigation as for the investigation of the formation of D_{n+2}, it was necessary to make p3 rate limiting. This was accomplished by selecting two parameters: (1) The concentration of the deoxynucleoside triphosphate used to produce D_{n+2} was set at more than 10 times the experimentally determined K_{2map} value. (2) The concentrations of dATP selected gave plots of the right side of eq 1 against t, with f_{n+3} substituted for f_{n+2} and $r = 0$, that were independent of time. Since p3 is rate limiting, the mathematical formalism is the same as the previous case for p2. After a survey experiment to determine the concentration range of dATP that satisfied the condition of making p3 rate limiting, an 8-fold concentration range of dATP was chosen, with the maximum concentration being about K_{3map}. The experimental dependence of p3 on dATP concentration was

$$p3 = \frac{V_{3map}}{1 + K_{3map}/[dATP]}$$

Table IV contains the results of these measurements. The sequence of additions of deoxynucleotides for D_{n+1}, D_{n+2}, and D_{n+3} was dTTP, dNTP, and dATP.

Misinsertion Frequency and Stable Misinsertion Frequency. The mathematical model developed in Appendix A gives the ratio of the apparent V_{max} and K_m, V_{2map}/K_{2map}, as $V_{2m}/(K_{2D}K_{2m})$, eq 8, where V_{2m} and K_{2m} are the true values for the addition of deoxynucleotide and K_{2D} is the dissociation constant of the template–primer D_{n+1}. Creighton et al. (1992), Kuchta et al. (1988), and Wong et al. (1991) have shown that K_{2D} depends on the last base pair of the template–primer with K_{2D} being larger for mismatched base pairs than for standard base pairs. For all of the values of V_{2map} and K_{2map} in Table III the same base pair, T/A, is at the end of the template–primer. Thus the measure of the misinsertion frequency, $[V_{2map}(w)/K_{2map}(w)]/[V_{2map}(r)/K_{2map}(r)]$ (Boosalis et al., 1987), where w and r are the "wrong" and "right" nucleotides, will have the K_{2D} values cancel, giving the ratio in terms of the true V_{max}/K_m ratio.

The misinsertion frequency does not take into account the different rates of addition of the next nucleotide or the different rates of the 3′–5′ exonuclease activity. In Appendix B, an expression for the relative rates of incorporation of the next standard nucleotide is given in terms of the parameters p2, pn2, and p3, eqs B1 and B2. The ratio of the relative rate of the incorporation of a standard nucleotide after the incorporation of a "wrong" nucleotide to the relative rate of incorporation of the standard nucleotide after the incorporation of the "right" nucleotide is the stable misinsertion frequency. Following the definition of the misinsertion frequency, p2 is taken to be V_{2map}/K_{2map}. Because p3 appears in the expression (p3 + pn2), it is necessary to calculate p3 at some concentration of dN_3TP so that the dimensions of p3 will agree with the dimensions of pn2, namely, t^{-1}. Taking p3 equal to V_{3map} computes the maximum stable misinsertion frequency because it minimizes the exonuclease contribution, as discussed in Appendix B. As an illustration of the calculation, the insertion of A opposite T^h is compared to the insertion of G^s opposite T^h. From Table III from template A-T^h-T-T, p2(A) = 1.3×10^{-3}, p2(G^s) = 1.5, pn2(A) = 3.4×10^{-2}, and pn2(G^s) = 3.0×10^{-2}. From Table IV for the template A-T^h-T-T, p3-(A,A) = 4.8×10^{-3}, and p3(G^s,A) = 7.0×10^{-3}. From eqs B1 and B2, Appendix B, f(A) = 5.0×10^{-7}, and f(G^s) = 1.3×10^{-3}. The maximum stable misinsertion frequency is F_s = 4×10^{-4}.

Replication of G^s/T^h. The kinetics parameters for the replication of the base pairs G/C and G^s/T^h are given in Table III. There are three significant differences between the two base pairs: (1) The V_{2map} values of the G/C base pair are greater than the V_{2map} values of the G^s/T^h base pair. (2) The K_{2map} values of the G^s/T^h base pair are lower than the those of the G/C base pair. (3) The 3′–5′ exonuclease rate, pn2, is significant for the G^s/T^h base pair and not detectable for the G/C base pair. Although the V_{2map} and K_{2map} values are considerably different for the two base pairs, both pairs have the same range of values of the insertion efficiency, V_{2map}/K_{2map}.

Fidelity of G^s/T^h with A/T. The kinetics parameters for the insertion of A and T opposite T^h are given in Table III. The insertion efficiency of T is very low, but the insertion efficiency of A is moderately high, 1.3×10^{-3}. The insertion efficiencies for A and T opposite G^s are 9.0×10^{-4} and 5.0×10^{-4}. The kinetics parameters for the insertion of G^s opposite T were not determined.

Fidelity of G^s/T^h with G/C. The insertion efficiency of T^h opposite G is less than 10^{-3}. Because of the very small amount of T^h inserted across from G, it was not possible to determine the rate of removal of T^h by the exonuclease. The insertion of G opposite G^s was too low to detect even at 120 min.

Extension of G^s/T^h and G/C Base Pairs by dATP. An important determination with a new base pair is whether replication can take place beyond its position in a template–primer at an acceptable rate. In addition, the rate of replication beyond a mismatched base pair compared to the 3′–5′ exonuclease activity at the mismatch determines the efficiency of proofreading. To allow the mathematical model developed in Appendix A to be used to evaluate these rates of replication, it is necessary to make step p3, Scheme II, rate limiting. The general methods to accomplish this rate limitation were discussed in the Results under V_{3map} and K_{3map} from an

Table III: Summary of the Kinetics Parameters for Nucleotide Insertion and Removal by the Large Klenow Fragment of *Escherichia coli* DNA Polymerase I

no.	insertion	template,[a] 3′–5′	T/P[b]	V_{2map},[c] min^{-1}	K_{2map},[c] μM	pn2,[d] min^{-1}	V_{2map}/K_{2map}, min^{-1} μM^{-1}	no. of expts[e]
1	G^s	-A-*T^h*-T-T	1/1	7.4×10^{-2} [f]	5.0×10^{-2} [f]	3.0×10^{-2} [f]	1.5	3
2	G^s	-A-*C*-T-G	5/1	1.1×10^{-1} [g]	3.0×10^{-1} [g]	nd[h]	3.7×10^{-1}	2
3	T^h	-A-A-*G^s*-T-	3/2	1.7×10^{-2} [f]	2.5×10^{-2}	6.0×10^{-2} [f]	7.0×10^{-1}	3
4	T^h	-A-*G*-C-T	4/1	$<6.0 \times 10^{-3}$	$>1.3 \times 10$	nd[h]	$<5.0 \times 10^{-4}$	2
5	G	-A-*C*-T-G	5/1	1.3×10^{-1} [g]	8.9×10^{-2} [g]	nd[h]	1.5	5
6	G	-A-*T^h*-T-T	1/1	7.4×10^{-2} [f]	1.8×10^{-1} [f]	5.0×10^{-2} [f]	4.0×10^{-1}	2
7	G	-A-A-*G^s*-T-	3/2	nd[h]	nd[h]	nd[h]		
8	C	-A-*G*-C-T	4/1	1.3×10^{-1} [g]	2.5×10^{-1} [g]	nd[h]	5.0×10^{-1}	2
9	C	-A-A-*G^s*-T-	3/2	5.0×10^{-2} [f]	2.5×10^{-1} [f]	2.0×10^{-2} [f]	2.0×10^{-1}	2
10	C	-A-*T^h*-T-T	1/1	4.0×10^{-4} [f]	3.0[f]	nd[h]	1.3×10^{-4}	2
11	A	-A-*T^h*-T-T	1/1	4.9×10^{-2} [f]	3.7×10[f]	3.4×10^{-2} [f]	1.3×10^{-3}	2
12	A	-A-A-*G^s*-T-	3/2	5.6×10^{-3} [f]	6.0[f]	nd[h]	9.0×10^{-4}	2
13	T	-A-*T^h*-T-T	1/1	1.1×10^{-4} [f]	1.6×10[f]	nd[h]	7.0×10^{-6}	2
14	T	-A-A-*G^s*-T-	3/2	2.0×10^{-2} [f]	4.0×10[f]	nd[h]	5.0×10^{-4}	2

[a] The template base opposite the inserted base is italicized. [b] The template/primer used; see Table I. [c] Calculated by the procedures presented in the Results section under Determination of V_{2map}, K_{2map}, and pn2. [d] The 3′–5′ exonuclease rate calculated by the procedures in the Results section under Determination of V_{2map}, K_{2map}, and pn2. [e] Number of independent experiments. [f] The standard error of the mean is 0.33. [g] The standard error of the mean is 0.2. [h] Not detected.

Table IV: Summary of the Kinetics Parameters for Extension of the Primer by dATP after a Base-Pair Insertion from Table III

no.	insertion[a]	template, 3′–5′	T/P[b]	V_{3map},[c] min^{-1} [d]	K_{3map},[c] μM[d]	V_{3map}/K_{3map}, min^{-1} μM^{-1}	no. of expts
1	G^s,A	-A-*T^h*-*T*-T	1/1	7.0×10^{-3}	7.8×10^{-2}	9.0×10^{-2}	2
2	G^s,A	-A-*C*-*T*-G	5/1	6.0×10^{-3}	1.4	4.0×10^{-3}	2
3	T^h,A	-A-A-*G^s*-*T*-	3/2	1.2×10^{-2}	1.7×10^{-1}	1.2×10^{-1}	2
4	G,A	-A-*C*-*T*-G	5/1	4.5×10^{-2}	1.0	4.5×10^{-2}	2
5	G,A	-A-*T^h*-*T*-T	1/1	2.5×10^{-2}	1.1	2.3×10^{-2}	2
6	A,A	-A-*T^h*-*T*-T	1/1	4.8×10^{-3}	nm[e]	nm[e]	2
7	C,A	-A-A-*G^s*-*T*-	3/2	2.5×10^{-2}	1.7×10^{-1}	1.5×10^{-1}	2
8	A,A	-A-A-*G^s*-*T*-	3/2	nd[f]	nd[f]		2

[a] The first base is inserted opposite the first italicized base in the template. The second base, A, is inserted opposite the second italicized base in the template. All base insertions before the italicized bases in the template are T. [b] The template/primer used; see Table I. [c] Calculated by the procedures presented in the Results section under V_{3map} and K_{3map} from an Extension to D_{n+3} by dATP. [d] Standard error of the mean is 0.33. [e] Not measured. [f] Not detected.

Extension to D_{n+3}. One condition to make p3 rate limiting was to saturate the polymerase with dN_2TP, step p2, Scheme II. Two examples: (1) K_{2map} for T^h/G^s is 2.5×10^{-2}, and the concentration of dT^hTP was 2 μM; (2) K_{2map} for G/C is 8.9×10^{-2}, and the concentration of dGTP was 2 μM. Table IV lists the results of these measurements. The sequence of bases added is given in the second column, and the sequence of pairing bases in the template is shown in italics in the third column. For example, in line 1, G^s is added across from T^h in the template, and then A is added across from T in the template. The value of the insertion efficiency, V_{3map}/K_{3map}, after G^s/T^h is about the same as after T^h/G^s. The insertion efficiency after G^s/T^h is a factor of 2 greater than after the G/C base pair. There was no detectable elongation after the A/G^s base pair at 120 min.

Model Selection and Data Consistency. The V_{2map} for kinetics model I is $k_4[E_{2T}]$; for model II, $k_4k_6[E_{2T}]/k_{-4}$. Model I predicts that all of the V_{2map} values should be equal for each type of template–primer. The variation in V_{2map} in Table III suggests that model II is correct. However, excluding lines 10, 12, and 13, all of the values are just on the border of experimental error. The assays for lines 10, 12, and 13 used very high concentrations of nucleotides. It is possible that at these concentrations a significant amount of polymerase combined with nucleotide, thus reducing the amount of free enzyme. This would invalidate an essential assumption in the derivation of the models, namely, that the free enzyme concentration is essentially the total enzyme concentration. Another possibility is nonproductive binding of the dNTPs with extreme mismatching, for example, a Hoogsteen-type pairing for A/G^s in line 12.

DISCUSSION

The goal of this investigation was to determine the following: (1) Does the base pair G^s/T^h replicate in vitro with KF at an acceptable rate? (2) Can the base pair G^s/T^h replace the G/C base pair in replication by KF with acceptable fidelity? (3) Is there sufficient fidelity of replication of G^s/T^h and the standard base pairs so that three base pairs are possible? (4) What modifications of the base pair G^s/T^h would lead to greater fidelity?

Fresht et al. (1982) provided the idea that a useful parameter for evaluating the addition of a deoxynucleoside monophosphate to a primer opposite a particular base in a template by a polymerase is V_{max}/K_m, the insertion efficiency. A comparison of the kinetics parameters in Table III for the base pairs G^s/T^h and G/C shows that the insertion efficiencies, V_{2map}/K_{2map}, for both base pairs fall within the same range of values. A significant difference between the base pairs is the higher rate of removal of G^s and T^h by the 3′–5′ exonuclease, pn2. The limit of detection of the exonuclease rate is about $0.2V_{max}$ with the present data, so that the turnover rate for the base pair G/C of 6% reported by Fresht et al. (1982) would not have been observed.

With the new base pair the determination of the insertion efficiency of the next base after formation of the G^s/T^h base pair is important. Table IV shows that the insertion efficiency of A opposite T after a G^s/T^h base pair is about a factor of 2 greater than the insertion frequency of A opposite T after a G/C base pair. Mandelman et al. (1989) have shown that V_{max} and K_m can vary by more than 1 order of magnitude with the preceding base pair. The V_{2map} and K_{2map} values

determined in the present study are only for the preceding base pair T/A. Thus the results only suggest that the replication rate of the G^s/T^h base pair and the rate of addition of a standard base after G^s/T^h appear to be satisfactory.

Boosalis et al. (1987) have suggested that the appropriate measure of misinsertion is the ratio of the "incorrect" nucleotide insertion efficiency to the "correct" nucleotide insertion efficiency, designated the misinsertion frequency. For the base pair G^s/T^h to replace the base pair G/C, the discrimination against the incorporation of A or T across from G^s or T^h should be high. For reference, polymerase I has misinsertion frequencies of 8×10^{-5} for G opposite T, 2×10^{-5} for C opposite A, and 1×10^{-5} for G opposite A (Fresht et al., 1982). Replication by KF without the 3′–5′ exonuclease has misinsertion frequencies of 6.5×10^{-4} for G opposite T and 7.5×10^{-5} for A opposite G (Bebenk et al., 1990). The values of the misinsertion frequencies of G opposite T for two other polymerases without an exonuclease range from 5.4×10^{-4} to 1.9×10^{-3} (Mendelman et al., 1989). The misinsertion frequency of T opposite G^s, Table III, is 7×10^{-4}, which is essentially the value of G opposite T for KF without an exonuclease. The misinsertion frequency of A opposite G^s is 1.3×10^{-3}. This is a factor of 10^2 greater than that of G opposite A for polymerase I and a factor of 10 greater than that of KF without exonuclease. However, when the rate of addition of the next base, A opposite T, was examined (Table IV), the rate was so low it could not be detected. The misinsertion frequency of A opposite T^h is 9×10^{-4}. Since there is a significant exonuclease rate for removing A opposite T^h, it is useful to define a stable misinsertion frequency, Appendix B, which makes a quantitative estimate of the effect of the exonuclease and the rate of insertion of the next standard base on fixing a "wrong" base compared to a "right" base. The actual value of the stable misinsertion frequency depends on the concentration of the standard nucleotide used to extend the primer. The largest value is obtained if the contribution of the exonuclease activity is minimized by taking the concentration of the nucleotide to be saturating. For this condition the stable misinsertion frequency is called the maximum. When the exonuclease activity, Table III, and the addition of the next base, A opposite T, Table IV, is taken into account, the maximum stable misinsertion frequency for A opposite T^h is 4×10^{-4}. This is a factor of 5 greater than the misinsertion frequency of polymerase I for G opposite T. Because only one base pair, A/T, was used for the extension, the results only suggest that the replication of the G^s/T^h base pair by KF has about the same fidelity as the G/C base pair when the stable misinsertion frequencies are used.

The final set of misinsertions to consider are the combinations between G^s, T^h, G, and C. The misinsertion efficiency of T^h opposite G is $<1 \times 10^{-3}$, Table III. However, the reciporal misinsertion efficiency of G opposite T^h is 0.27. If the exonuclease activity and the addition of the next base, A, Table IV, are taken into account, the maximum stable misinsertion efficiency is 4×10^{-2}. No exonuclease activity was detected when G^s was inserted opposite C, although exonuclease activity was detected for the insertion of C opposite G^s. A reason for the apparent discrepancy is that if the same rate of exonuclease activity was present for the insertion of G^s opposite C as for C opposite G^s, its detection would be within the noise level of the experiment. The misinsertion frequency of G^s opposite C is 0.25. If the maximum stable misinsertion frequency of G^s opposite C is calculated without any exonuclease activity, then the result is 3×10^{-2}. If the exonuclease rate for C opposite G^s is used, the value is 7×10^{-3}. The maximum stable misinsertion efficiency of C opposite G^s is 0.26. The results show that the moderate levels of fidelity for the insertion of G^s opposite C, C opposite G^s, and G opposite T^h preclude the use of the base pair G^s/T^h as a third base pair for high-fidelity replication by KF. A summary of the misinsertion frequencies and the maximum stable misinsertion frequencies is given in Table V. The large discrimination against T^h insertion opposite G was anticipated many years ago by Laland and Serck-Hanssen (1964). More recently, a brief communication indicated that the insertion of G opposite 2-pyrimidinone was low (Charczuk et al., 1986).

Table V: Misinsertion Frequencies and Maximum Stable Misinsertion Frequencies[a]

insertion/ template	misinsertion frequency[b]	max stable misinsertion frequency[c]
T/G^s	7.0×10^{-4}	nm[d]
A/G^s	1.3×10^{-3}	0 (no elongation)
A/T^h	9.0×10^{-4}	4.0×10^{-4}
T/T^h	4.0×10^{-6}	nm[d]
T^h/G	$<1 \times 10^{-3}$	nm[d]
G/T^h	3.0×10^{-1}	6.0×10^{-2}
G^s/C	3.0×10^{-1}	3×10^{-2}–7×10^{-3}
C/G^s	3.0×10^{-1}	2.6×10^{-1}

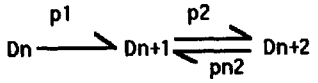
[a] The data for the calculations is taken from Tables III and IV. A detailed calculation is given in the Results section under Misinsertion Frequency and Stable Misinsertion Frequency. [b] Calculated according to Mendelman et al. (1989). [c] As defined in Appendex B. [d] Not measured.

Ling et al. (1992) have published an investigation of the incorporation of G^s in place of G or A using KF and bacteriophage M13 DNA as the template. Because the conditions used for the assay are different from the conditions in the present investigation, precise comparisons may be misleading. If a simple correspondence is assumed, two comparisons of the results can be made: (1) Ling et al. found a misinsertion frequency of G^s opposite T of 1.35×10^{-3}. The misinsertion frequency of T opposite G^s in the present investigation is 7×10^{-4}. (2) Ling et al. found a misinsertion frequency of G^s opposite C of 0.19 compared to the value of 0.3 in the present investigation. Considering that Ling's results are averages for different preceding and following standard base pairs, the numerical agreement is remarkable.

Is there any biological significance to the parameters determined in the present investigation, where the concentration of the template–primer was chosen so that the rate of each polymerization step was proportional to the concentration of the template–primer species used? In a complete enzymological investigation of an ordered bisubstrate reaction, the true V_{max} and K_m values of the dNTPs would be determined from initial rates and the variation in the apparent values with the concentration of the template–primer. In a simplistic model of *E. coli* where the cell is a bag, the concentration of replication sites is on the order of 10^{-9}–10^{-10} M depending on how many repair sites are present. From the determinations of the present investigation, a concentration in this range makes the $1/[D_n]$ term in velocity equation A4, Appendix A, the dominant factor. Thus the same condition used in the present study is the condition in this model of the cell and suggests that the apparent values, V_{2map} and K_{2map}, are the biologically significant values.

A previous investigation (Rappaport, 1988) of the relative association constants of the various base-pair combinations showed the stability of the base pair G/C compared to the base pair G^s/C to be a factor of more than 150. In the present investigation the V_{2map}/K_{2map} values differed by a factor of 4. In addition, the normal values of the kinetic parameters

Scheme III



of the insertion of A opposite T after the insertion of C opposite G^s, Table IV, strongly indicates that very little distortion has been introduced. Model building and molecular energy calculations show a very large steric interference between the thio group of 6-thioguanine and the amino group of cytosine if normal Watson–Crick geometry is maintained. In an attempt to understand the significant misinsertion frequencies of G^s and C with each other, a molecular mechanics investigation was performed. The amber parameter set was used (Weiner et al., 1986). The conformation of 6-thioguanine was taken from an X-ray analysis (Thewalt & Bugg, 1972). Sodium ions were the counter ions. As a control to the utility of the investigation, the conformation of the base pair G/T in a duplex oligomer was determined. The conformation of the local energy minimum, starting from a canonical B-DNA structure, reproduced the conformation determined by X-ray analysis (Kneale et al., 1985). The general result of the investigation of the G^s/C base pair, regardless of the neighboring base pairs, was that two hydrogen bonds of standard length were formed between G^s and C. However, various degrees of propeller twist, vertical displacement, and buckling occurred between the two bases. Sponer and Kypr (1990) have given an analysis of the elimination of interstrand purine clash by base-pair buckling. The buckling conformation essentially damped out after one base pair on either side of the G^s/C base pair. The C1′ distances did not depart from the standard range of values. For the sequence used in this investigation, T–G^s–A, the plane of G^s remained perpendicular to the helix axis, C showed a propeller twist, and the plane of C was buckled with respect to G^s. If the validity of the molecular mechanics analysis is assumed, a possible solution to the moderate fidelity problem with the G^s/C base pair is to remove the amino group of 6-thioguanine to give 6-thiohypoxanthine. This solution would require the use of 2-thiopyrimidinone as the complementary pyrimidine. The use of a 2-thiopyrimidinone would be expected to significantly decrease the misinsertion efficiency of guanine opposite 2-thiopyrimidinone compared to guanine opposite 2-pyrimidinone.

ACKNOWLEDGMENT

I thank J. K. Derial for the synthesis of many standard oligomers, Barbara Brownstein for continuing intellectual interest and critical administrative support over many years, and Bill Brinigar for many significant technical discussions. The very constructive comments by a reviewer about the kinetics are acknowledged.

APPENDIX A

The formal set of reactions (Markov chain) to solve as a function of time is shown in Scheme III, where D_n is the template–primer complex with a primer of n nucleotides, p1 is the rate of conversion of D_n to D_{n+1}, p2 is the rate of conversion of D_{n+1} to D_{n+2}, and pn2 is the rate of decay of D_{n+2} to D_{n+1}. The decays of D_{n+1} to D_n and of D_n to D_{n-1} could be included. However, since the experimental conditions of interest in the present investigation will allow the neglect of their contributions, the added complexity did not seem useful. The rate equations are

$$d[D_n]/dt = (-p1)[D_n],$$

$$d[D_{n+1}]/dt = (p1)[D_n] - (p2)[D_{n+1}] + (pn2)[D_{n+2}]$$

$$d[D_{n+2}]/dt = (p2)[D_{n+1}] - (pn2)[D_{n+2}]$$

Let $f_n(t)$ represent the fraction of the total template–primer in the form D_n at time t; the solutions to the rate equations (Zill & Cullen, 1992) are

$$f_n(t) = \exp[(-p1)t] \tag{A1a}$$

$$f_{n+1}(t) = \frac{pn2}{p2 + pn2} + \frac{(p1)(p2)\exp[-(p2 + pn2)t]}{p_1} - \frac{(pn2 - 1)\exp[(-p1)t]}{p_2} \tag{A1b}$$

$$f_{n+2}(t) = \frac{p2}{p2 + pn2} - \frac{(p1)(p2)\exp[-(p2 + pn2)t]}{p_1} + \frac{(p2)\exp[(-p1)t]}{p_2} \tag{A1c}$$

$$p_1 = (p1)(pn2) + (p2)(p1) - (pn2)^2 - 2(p2)(pn2) - (p2)^2$$

$$p_2 = p1 - p2 - pn2$$

For the case where p1 $\gg$ p2 or pn2, that is, p2 is the rate-limiting step, the fraction of template–primer in the form D_{n+2}, $f_{n+2}(t)$, at time t is

$$f_{n+2}(t) = \frac{(p2)\,[1 - \exp[-(p2 + pn2)t]]}{p2 + pn2} \tag{A2}$$

which with the substitution of $r = pn2/p2$ and rearrangement becomes

$$\frac{-\ln[1 - (1 + r)f_{n+2}(t)]}{(1 + r)t} = p2 \tag{A3}$$

For the present investigation the minimum kinetics formalism is as follows:

step 1: $E + D_n \leftrightarrow E(D_n)$ (k_1, k_{-1})

step 2: $E(D_n) + N_1TP \leftrightarrow E(D_n,N_1TP)$ (k_2, k_{-2})

step 3: $E(D_n,N_1TP) \rightarrow E(D_{n+1}) + pp_i$ (k_3)

step 4: $E + D_{n+1} \leftrightarrow E(D_{n+1})$ (k_4, k_{-4})

step 5: $E(D_{n+1}) + N_2TP \leftrightarrow E(D_{n+1},N_2TP)$ (k_5, k_{-5})

step 6: $E(D_{n+1},N_2TP) \rightarrow E(D_{n+2}) + pp_i$ (k_6)

step 7: $E + D_{n+2} \leftrightarrow E(D_{n+2})$ (k_7, k_{-7})

step 8: $E(D_{n+2}) \rightarrow E(D_{n+1})$ (k_8)

where E is the polymerase, D_n is the template–primer, with the primer having a length of n nucleotides, N_iTP is a deoxynucleoside triphosphate, the k values are rate constants, and pp_i is pyrophosphate.

The following three assumptions are made: (1) The concentration of pyrophosphate is low enough so that steps 3 and 6 are irreversible. (2) The rate of change of the concentrations of the intermediates is sufficiently slow in time so that the steady-state expression is a good approximation as the concentrations of D_n, D_{n+1}, and D_{n+2} change. (3) The

total concentration of enzyme–substrate intermediates is small compared to the total enzyme concentration.

The steady-state solution of reactions 1–3 is

$$-\frac{d[D_n]}{dt} = \frac{V_{1m}}{1 + \frac{K_{1m}}{[N_1TP]} + \frac{1}{[D_n]}\left(K_1 + \frac{K_{1m}K_{1D}}{[N_1TP]}\right)} \quad (A4)$$

where

$$[E_{1T}] = [E] + [E(D_n)] + [E(D_n,N_1TP)] = [E_T] - [E(D_{n+1})] - [E(D_{n+1},N_2TP)] - [E(D_{n+2})]$$

$$V_{1m} = k_3 \times [E_{1T}]$$

$$K_{1m} = (k_3 + k_{-2})/k_2$$

$$K_1 = k_3/k_1$$

$$K_{1D} = k_{-1}/k_1$$

The enzyme is regenerated in step 4. The equation for step 4 has been written backward from the usual convention for product formation to conform with step 1. Because step 3 is irreversible and because of assumption 3, it is not necessary to explicitly include $E(D_{n+1})$ in the derivation.

Following Alberty and Koerber (1957), the steady-state expression is integrated over time from 0 to t. The resulting expression is

$$([D_n(t)] - [D_n(0)])\left(1 + \frac{K_{1m}}{[N_1TP]}\right) + \left(K_1 + \frac{K_{1D}K_{1m}}{[N_1TP]}\right)\ln\frac{[D_n(t)]}{[D_n(0)]} = -V_{1m}t \quad (A5)$$

When the logarithmic term dominates, eq A1a and A5 are equivalent, so that p1 is given by

$$p1 = \frac{V_{1m}}{K_1\left(1 + \frac{K_{1D}K_{1m}}{K_1[N_1TP]}\right)}$$

The apparent V_{max} (V_{1map}) and the apparent K_m (K_{1map}) are

$$V_{1map} = \frac{V_{1m}}{K_1} = k_1[E_{1T}] \qquad K_{1map} = \frac{K_{1m}K_{1D}}{K_1}$$

The logarithmic term is dominant when the value of $[D_n]$ in eq A4 is sufficiently small that only terms in the denominator with $1/[D_n]$ are important.

Kinetics Model I

If $k_4[E][D_{n+1}]$ is much greater than $k_3[E(D_n,N_1TP]$, the steady-state expression for reactions 4–6 is

$$\frac{d[D_{n+2}]}{dt} = \frac{V_{2m}}{1 + \frac{K_{2m}}{[N_2TP]} + \frac{1}{[D_{n+1}]}\left(K_2 + \frac{K_{2m}K_{2D}}{[N_2TP]}\right)} \quad (A6)$$

where

$$[E_{2T}] = [E] + [E(D_{n+1})] + [E(D_{n+1},N_2TP)] = [E_T] - [E(D_n)] - [E(D_n,N_1TP)] - [E(D_{n+2})]$$

$$V_{2m} = k_6[E_{2T}]$$

$$K_{2m} = (k_6 + k_{-5})/k_5$$

$$K_2 = k_6/k_4$$

$$K_{2D} = k_{-4}/k_4$$

The enzyme is regenerated in step 7. The equation for step 7 is written backward from the normal convention for product formation to conform to step 1. Because step 6 is irreversible and because of assumption 3, it is not necessary to explicitly include $E(D_{n+2})$ in the derivation. When $[D_{n+1}]$ is sufficiently small, only the terms in the denominator with $1/[D_{n+1}]$ are important and p2, eq A3, is given by

$$p2 = \frac{V_{2m}}{K_2\left(1 + \frac{K_{2m}K_{2D}}{K_2[N_2TP]}\right)} \quad (A7)$$

The apparent V_{max} (V_{2map}) and the apparent K_m (K_{2map}) are

$$V_{2map} = \frac{V_{2m}}{K_2} = k_4[E_{2T}] \qquad K_{2map} = K_{2m}K_{2D}/K_2$$

and V_{2map}/K_{2map} is given by

$$V_{2map}/K_{2map} = V_{2m}/K_{2D}K_{2m} \quad (A8)$$

Kinetics Model II

If an equilibrium is assumed between D_{n+1}, E, and $E(D_{n+1})$ and between D_{n+1}, $E(N_2TP)$, and $E(D_{n+1},N_2TP)$, that is, $k_{-4} \gg k_5[N_2TP]$, then

$$[E(D_{n+1})] + [E(D_{n+1},N_2TP)] = \frac{[E][D_{n+1}]}{K_{2D}} + \frac{[E(N_2TP)][D_{n+1}]}{K_{2D'}}$$

If the steady-state expression derived from steps 5 and 6 of the kinetics formalism,

$$k_5[E(D_{n+1})][N_2TP] = (k_{-5} + k_6)[E(D_{n+1},N_2TP)]$$

and the condition $[E(N_2TP)] \ll [E]$, obtained by the choice of the concentration range of N_2TP employed and verified by the experimental dependence of the velocity on $[N_2TP]$, are used, then p2 is given by

$$p2 = \frac{V_{2m}}{K_{2D}(1 + K_{2m}/[N_2TP])} \quad (A9)$$

where V_{2m} and K_{2m} are as defined in eq A6, and $V_{2map} = V_{2m}/K_{2D} = (k_4/k_{-4})k_6[E_{2T}]$. Thus V_{2map}/K_{2map} is the same as in eq A8.

When the concentration of D_{n+2} is low, the steady-state expression for the exonuclease rate identifies pn2 as

$$pn2 = k_8[E_{3T}]/K_{3m} \quad (A10)$$

where

$$[E_{3T}] = [E] + [E(D_{n+2})] = [E_T] - [E(D_n)] - [E(D_n,N_1TP)] - [E(D_{n+1})] - [E(D_{n+1},N_2TP)]$$

$$K_{3m} = (k_{-7} + k_8)/k_7$$

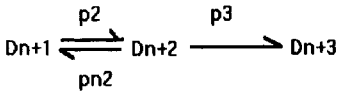
In the derivations of eq A6–A10, $[E_{1T}]$, $[E_{2T}]$, and $[E_{3T}]$ have been taken as constants. They are defined as follows:

$$[E_{1T}] = [E] + [E(D_n)] + [E(D_n,N_1TP)]$$

$$[E_{2T}] = [E] + [E(D_{n+1})] + [E(D_{n+1},N_2TP)]$$

$$[E_{3T}] = [E] + [E(D_{n+2})] \quad (A11)$$

Scheme IV



Because the concentrations of the intermediates change in time, the only way p1, p2, and pn2 can remain constant in time is when [E] dominates each of the sums. In this case, [E] is essentially $[E_T]$. This is the reason it is necessary for assumption 3 to be satisfied by using a sufficiently low concentration of template–primer.

APPENDIX B

Stable Misinsertion Frequency and Maximum Stable Misinsertion Frequency. The formal steps to be considered are shown in Scheme IV, using the same notation as in Scheme III. The goal is to calculate the rate of arrival at D_{n+3} when there is a choice of a right (r) and a wrong (w) nucleotide for D_{n+1} going to D_{n+2}. The concentrations of the right and the wrong deoxynucleotide triphosphates will be taken as being equal and so low that the rates are proportional to the concentrations (Mendelman et al., 1989). If a particular template–primer is followed at D_{n+1}, then the probability that a wrong nucleotide is incorporated in going to D_{n+2} is

$$\mathrm{p2(w)/p2(w) + p2(r)}$$

Once D_{n+2} has been formed, the probability that D_{n+2}(w) will go to D_{n+3} is

$$\frac{\mathrm{p3(w)}}{\mathrm{p3(w) + pn2(w)}}$$

and the rate is p3(w). If the probabilities and rates are combined, the relative rate of D_{n+1} going to D_{n+3} with a wrong nucleotide is

$$f(\mathrm{w}) = \frac{[\mathrm{p2(w)}][\mathrm{p3(w)}][\mathrm{p3(w)}]}{[\mathrm{p2(w) + p2(r)}][\mathrm{p3(w) + pn2(w)}]} \quad \text{(B1)}$$

For the incorporation of the right nucleotide the result is

$$f(\mathrm{r}) = \frac{\mathrm{p2(r)[p3(r)][p3(r)]}}{[\mathrm{p2(w) + p2(r)}][\mathrm{p3(r) + pn2(r)}]} \quad \text{(B2)}$$

The stable misinsertion frequency, F_s, is defined as the ratio of $f(\mathrm{w})/f(\mathrm{r})$. In analogy with Mendelman et al. (1989), p2 is taken as $V_{2\mathrm{map}}/K_{2\mathrm{map}}$. Because p3 appears in the expression (p3 + pn2), it is necessary to assign a value to $[dN_3TP]$ to establish a value of p3 that has the same dimensionality as pn2, namely, t^{-1}. One choice would be the value of $[dN_3TP]$ that makes F_s a minimum. Since, generally, pn2(w) > pn2(r), it is useful to make an evaluation that states the worst case. For this purpose p3 is taken to be $V_{3\mathrm{map}}$. This causes the exonuclease rate, pn2, to have the least effect. Physically this corresponds to the maximum trapping of the wrong nucleotide. The stable misinsertion frequency evaluated with p3 = $V_{3\mathrm{map}}$ is called the maximum stable misinsertion frequency.

REFERENCES

Alberty, R. A., & Koerber, B. M. (1957) *J. Am. Chem. Soc. 79*, 6379.

Bebenek, K., Joyce, C. M., Fitzgerald, M. P., & Kunkel, T. A. (1990) *J. Biol. Chem. 265*, 13878.

Boosalis, M. S., Petruska, J., & Goodman, M. F. (1987) *J. Biol. Chem. 262*, 14689.

Charczuk, R., Tamm, C., Suri, B., & Bickle, T. A. (1986) *Nucleic Acids Res. 14*, 9530.

Cleland, W. W. (1967) *Adv. Enzymol. Relat. Areas Mol. Biol. 29*, 1.

Creighton, S., Huang, M., Cai, H., Arnheim, N., & Goodman, M. F. (1992) *J. Biol. Chem. 267*, 2633.

Fersht, A. R., Knill-Jones, J. W., & Tsui, W. C. (1982) *J. Mol. Biol. 156*, 37.

Fox, J. J., Wempen, I., Hampton, A., & Doerr, I. G. (1958) *J. Am. Chem. Soc. 80*, 1669.

Freish, W., & Cramer, F. (1978) in *Nucleic Acid Chemistry* (Townsend, L. B., & Tipson, R. S., Eds.) pp 827–836, Wiley & Sons, New York.

Gildea, B., & McLaughlin, L. W. (1989) *Nucleic Acids Res. 17*, 2261.

Hoard, D. E., & Ott, D. G. (1965) *J. Am. Chem. Soc. 87*, 1785.

Kneale, G., Brown, T., & Kennard, O. (1985) *J. Mol. Biol. 186*, 805.

Kuchta, R. D., Benkovic, P., & Benkovic, S. J. (1988) *Biochemistry 27*, 6716.

Laland, G. G., & Serck-Hansen, G. (1964) *Biochem. J. 96*, 76.

Ling, Y., Nelson, J. A., Farquhai, D., & Beattie, K. L. (1992) *Nucleosides Nucleotides 11*, 23.

Matteucci, M. D., & Caruthers, M. H. (1981) *J. Am. Chem. Soc. 103*, 3185.

Maybaum, J., Bainnson, A. N., Roethel, W. M., Ajmera, S., Iwaniec, L. M., TerBush, D. R., & Kroll, J. (1987) *Mol. Pharmacol. 32*, 606.

McBride, L. J., & Caruthers, M. H. (1983) *Tetrahedron Lett. 24*, 5843.

Mendelman, L. V., Boosalis, M. S., Petruska, J., & Goodman, M. F. (1989) *J. Biol. Chem. 264*, 14415.

Michniewicz, J. J., Bahl, C. P., Itakura, K., Katagiri, N., & Narang, S. A. (1973) *J. Chromatogr. 85*, 159.

Piccirilli, J. A., Krauch, T., Moroney, S. E., & Benner, S. A. (1990) *Nature 343*, 33.

Rappaport, H. P. (1988) *Nucleic Acids Res. 16*, 7253.

Roark, D. N., Melin, D. H. V., & Jagow, R. H. (1978) in *Nucleic Acid Chemistry* (Townsend, L. B., & Tipson, R. S., Eds.) Part 2, p 583, Wiley & Sons, New York.

Sponer, J., & Jaroslav, K. (1990) *J. Biomol. Struct. Dyn. 7*, 1211.

Switzer, C. Y., Moroney, S. E., & Benner, S. A. (1989) *J. Am. Chem. Soc. 111*, 8322.

Tanaka, T., & Letsinger, R. L. (1982) *Nucleic Acids Res. 10*, 3249.

Thewalt, U., & Bugg, C. E. (1972) *J. Am. Chem. Soc. 94*, 8892.

Van Kampen, N. G. (1981) *Stochastic Processes in Physics and Chemistry*, Chapters IV and V, North–Holland, Amsterdam.

Weiner, S. J., Kollman, P. A., Nguyen, D. T., & Case, D. A. (1986) *J. Comput. Chem. 7*, 230.

Wong, I., Patel, S. S., & Johnson, K. A. (1991) *Biochemistry 30*, 526.

Yoshikawa, M., Kata, T., & Takenski, T. (1967) *Tetrahedron Lett.* 5065.

Zill, D. G., & Cullen, M. R. (1992) *Advanced Engineering Mathematics*, Chapter 4, PWS-Kent, Boston.