

Improving the fidelity of *Thermus thermophilus* DNA ligase

Jiaying Luo⁺, Donald E. Bergstrom¹ and Francis Barany*

Department of Microbiology, Box 62, Hearst Microbiology Research Center, Strang Cancer Prevention Center, Cornell University Medical College, 1300 York Avenue, New York, NY 10021, USA and ¹Department of Medicinal Chemistry and Pharmacognosy, Hansen Life Sciences Research Building, Purdue University, West Lafayette, IN 47907 and Walther Cancer Institute, Indianapolis, IN 46208, USA

Received February 15, 1996; Revised and Accepted May 28, 1996

ABSTRACT

The DNA ligase from *Thermus thermophilus* (*Tth* DNA ligase) seals single-strand breaks (nicks) in DNA duplex substrates. The specificity and thermostability of this enzyme are exploited in the ligase chain reaction (LCR) and ligase detection reaction (LDR) to distinguish single base mutations associated with genetic diseases. Herein, we describe a quantitative assay using fluorescently labeled substrates to study the fidelity of *Tth* DNA ligase. The enzyme exhibits significantly greater discrimination against all single base mismatches on the 3'-side of the nick in comparison with those on the 5'-side of the nick. Among all 12 possible single base pair mismatches on the 3'-side of the nick, only T-G and G-T mismatches generated a quantifiable level of ligation products after 23 h incubation. The high fidelity of *Tth* DNA ligase can be improved further by introducing a mismatched base or a universal nucleoside analog at the third position of the discriminating oligonucleotide. Finally, two mutant *Tth* DNA ligases, K294R and K294P, were found to have increased fidelity using this assay.

INTRODUCTION

DNA ligases catalyze the final essential step in DNA replication, DNA repair and genetic recombination (1). Deficient DNA ligase activity causes serious pathologies in host cells. Abnormal accumulation of large DNA replication intermediates in Blooms syndrome cells and accumulation of Okazaki fragments during DNA replication in the DNA ligase I mutant cell line 46BR have been documented (2). Furthermore, since DNA ligases are required to complete any DNA excision repair process, these enzymes must efficiently join nicked DNA duplexes which have been corrected to perfect complementarity, without resealing strand breaks containing damaged or mismatched bases. Thus, recovery from DNA damage depends not only on the specificity

of DNA polymerase and DNA repair enzymes, but also on the fidelity of DNA ligase. While a substantial amount of information has been accumulated on the fidelity of DNA polymerases (for reviews see 3–5), less is known about the fidelity of DNA ligases.

In vitro experiments using plasmid or synthetic oligonucleotide substrates reveal that T4 DNA ligase exhibits a relaxed specificity, sealing nicks with 3' or 5' apurinic or apyrimidinic (AP) sites (6), 1 nt gaps (6), 3' and 5' A-A or T-T mismatches (7), 5' G-T mismatches (8) or 3' C-A, C-T, T-G, T-T, T-C, A-C, G-G or G-T mismatches (9). The apparent fidelity of T4 DNA ligase may be improved in the presence of spermidine, high salt and very low ligase concentrations, where only T-G or G-T mismatch ligations were detected (7,9). DNA ligase from *Saccharomyces cerevisiae* discriminates 3'-hydroxyl and 5'-phosphate termini separated by a 1 nt gap and 3' A-G or T-G mismatches, however 5' A-C, T-C, C-A or G-A mismatches had very little effect on ligation efficiency (10). Mammalian DNA ligases I and III show different efficiencies in ligating 3' C-T, G-T and T-G mismatches (11). The vaccinia virus DNA ligase efficiently discriminates against 1 and 2 nt gaps and 3' G-A, A-A, G-G or A-G (purine-purine) mismatches, but easily seals 5' C-T, G-T, T-T, A-C, T-C, C-C, G-G, T-G or A-G mismatches as well as 3' C-A, C-T, G-T, T-T or T-G mismatches (12).

Our laboratory has cloned the thermostable *Thermus thermophilus* DNA ligase (*Tth* DNA ligase), which we and others have used in the ligase chain reaction (LCR) and ligase detection reaction (LDR) for detecting infectious agents and genetic diseases (13–27). The success of these and future disease detection assays, such as identifying tumor-associated mutations in an excess of normal DNA, depend on the exquisite fidelity of *Tth* DNA ligase. Quantitative studies of most mismatch ligations on the 3'-side of the nick revealed that *Tth* DNA ligase demonstrates far greater fidelity than that reported for T4 ligase (13). In the present report, we extend this study through development of a fluorescence assay using DNA substrates containing all possible single base mismatches on either the 3'- or 5'-side of the nick. We also describe methods to improve the fidelity of DNA ligases.

* To whom correspondence should be addressed

⁺Present address: Surgical Research Laboratory, Children's Hospital, Harvard Medical School, 300 Longwood Avenue, Boston, MA 02115, USA

MATERIALS AND METHODS

Synthesis of oligonucleotide primers

Oligonucleotide primers were synthesized using reagents and a model 394 automated DNA synthesizer from Applied Biosystems Division, Perkin-Elmer Corp. (Foster City, CA). Fluorescent label was attached to the 5'-end of oligonucleotides using 6-FAM (6-carboxyfluorescein) amidites or attached to a 3'-amino group (C3-CPG column; Glen Research, Sterling, VA) using NHS-FAM (*N*-hydroxysuccinimide ester of FAM; Applied Biosystems Division, Perkin-Elmer Corp.). A universal nucleotide analog, 1-(2'-deoxy- β -D-ribofuranosyl)-3-nitropyrrole, herein designated **Q**, was synthesized, transformed to the phosphoramidite and oligonucleotides synthesized as described (28). All oligonucleotides used in this study were purified by polyacrylamide gel electrophoresis with recovery of DNA from gel slices using C-18 Sep-Pak Cartridges from Waters Division, Millipore (Bedford, MA).

5'-Phosphorylation of oligonucleotide probes

One nanomole of gel-purified oligonucleotide was phosphorylated in a 25 μ l reaction containing 50 mM Tris-HCl, pH 7.6, 10 mM MgCl₂, 1 mM EDTA, 10 mM DTT, 1 mM ATP and 10 U T4 polynucleotide kinase (New England Biolabs, Beverly, MA) at 37°C for 45 min. The reaction was quenched by adding 0.5 μ l 0.5 M EDTA and the kinase was heat inactivated by incubation at 64°C for 20 min. The phosphorylated oligonucleotides were stored at -20°C in 5 μ l aliquots before use.

Purification of wild-type and mutant *Tth* DNA ligase

Wild-type *Tth* DNA ligase was purified from an *Escherichia coli* strain containing the *Tth* ligase gene under *phoA* promoter control as described (29), with some modifications. Briefly, cells were grown overnight at 30°C in low phosphate (inducing) medium, harvested, resuspended in lysis buffer (20 mM Tris-HCl, pH 8.5, 1 mM EDTA, 10 mM 2-mercaptoethanol and 0.15 mM PMSF) and sonicated (29). After removal of cellular debris, the supernatant was adjusted to 20 mM Tris-HCl, pH 8.5, 50 mM KCl, 10 mM MgCl₂, 0.5 mM EDTA, 1 mM DTT and 2 mM 2-mercaptoethanol, incubated at 65°C for 30 min and cleared by centrifugation at 4°C. The supernatant was diluted with an equal volume of 10 mM Tris-HCl, pH 7.6, 0.5 mM EDTA, filtered and loaded onto a 10 ml Red-Sepharose column (Pharmacia) equilibrated with 20 mM Tris-HCl, pH 8.5, 50 mM KCl, 1 mM EDTA and 20% glycerol. The protein was eluted with a 30 ml linear salt gradient of 50 mM-1 M KCl (30) using an FPLC apparatus from Pharmacia. *Tth* DNA ligase eluted between 0.4 and 0.8 M KCl and fractions containing pure *Tth* DNA ligase (seen as a doublet of adenylated and deadenylated forms on Coomassie brilliant blue stained 7.5% polyacrylamide-0.1% SDS polyacrylamide gels) were pooled. The enzyme was precipitated with an equal volume of saturated ammonium sulfate. The resultant precipitate was dissolved in 1.5 ml dH₂O and dialyzed at 4°C against 500 ml storage buffer containing 10 mM Tris-HCl, pH 8.5, 1 mM EDTA, 1 mM DTT, 0.2 mg/ml BSA and 50% glycerol. Protein concentration was determined by the Bradford method (31). About 4 mg *Tth* DNA ligase was obtained from a 450 ml culture. Mutant *Tth* DNA ligase was partially purified as described previously (32).

Fidelity assays of nick closure by *Tth* DNA ligase

Each reaction was performed in 40 μ l buffer containing 20 mM Tris-HCl, pH 7.6, 10 mM MgCl₂, 100 mM KCl, 10 mM DTT, 1 mM NAD⁺ and 12.5 nM (500 fmol) nicked DNA duplex substrates. DNA probes were denatured (94°C for 2 min), re-annealed (65°C for 2 min) and ligation initiated by the addition of 0.125 nM (5 fmol) *Tth* DNA ligase and carried out at 65°C. Five microliter aliquots were removed at 0, 2, 4, 6, 8 and 23 h and mixed with 18 μ l stop solution (83% formamide, 8.3 mM EDTA and 0.17% Blue Dextran). To 5 μ l of this mixture, 0.5 μ l ROX-1000, a fluorescently labeled in-lane size standard (Applied Biosystems Division, Perkin-Elmer Corp.) was added. Samples were denatured at 93°C for 2 min, rapidly chilled on ice prior to loading on an 8 M urea-10% polyacrylamide gel and electrophoresed at 1400 V (constant voltage) on a model 373A automated DNA Sequencer (Applied Biosystems Division, Perkin-Elmer Corp.). Electrophoresis conditions were modified as suggested by the manufacturer. The gel was polymerized in 1.2 \times TBE (54 mM Tris-borate, 1.2 mM EDTA, pH 8.0) and was pre-run before loading samples in a running buffer of 0.6 \times TBE (27 mM Tris-borate, 0.6 mM EDTA, pH 8.0) for 30 min with an electrode polarity opposite to the normal run with samples. After pre-run and sample loading, the gel was run in 0.6 \times TBE in the normal top to bottom direction for 2.5 h. Fluorescently labeled ligation products were analyzed and quantified using Genescan 672 version 1.2 software (Applied Biosystems Division, Perkin-Elmer Corp.) and the results were plotted using DeltaGraph Pro3 software (DeltaPoint Inc., Monterey, CA).

Measurement of initial rates of perfect match and mismatch ligations by *Tth* DNA ligase

Conditions for these experiments were the same as that for the fidelity assay except that different amounts of *Tth* DNA ligase and different probes were used, as indicated in the figure legends. Reactions were carried out in 160 μ l reaction buffer containing 12.5 nM (2 pmol) nicked DNA duplex substrates at 65°C. DNA probes and target were denatured by incubating the reaction mixture at 94°C for 2 min and re-annealed at 65°C for 2 min. Ligations were initiated by the addition of the indicated amount of the *Tth* DNA ligase. Aliquots (10 μ l) were removed at 0, 2, 4, 6, 8 and 10 h for reactions containing mismatched substrates and at 0, 10, 20, 30, 40, 50, 60 and 70 s for reactions containing matched substrates. For assays using matched substrates, 2.5 μ l samples were mixed with 2.5 μ l loading buffer and 0.5 μ l ROX-1000 before gel electrophoresis. Since the linear detection range of fluorescent samples on the 373A DNA Sequencer is from 0.1 to 10 fmol, products from mismatch ligation with a yield <1% were concentrated by ethanol precipitation for accurate quantification. From 10 μ l aliquots, 9.5 μ l were brought up to 200 μ l with TE buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA) and ethanol precipitated with 4 μ g yeast tRNA as carrier. The pellet was resuspended in 5 μ l loading buffer and 0.5 μ l ROX-1000 before gel electrophoresis. The amount of unreacted fluorescent probe was determined by diluting 0.5 μ l of the 10 μ l aliquot with 4.5 μ l loading buffer plus 0.5 μ l ROX-1000. Samples were separated on denaturing polyacrylamide gels and results analyzed as described above. The initial rates were calculated as the slope of the straight line in the graph with the x-axis as the time and the y-axis as the yield of products.

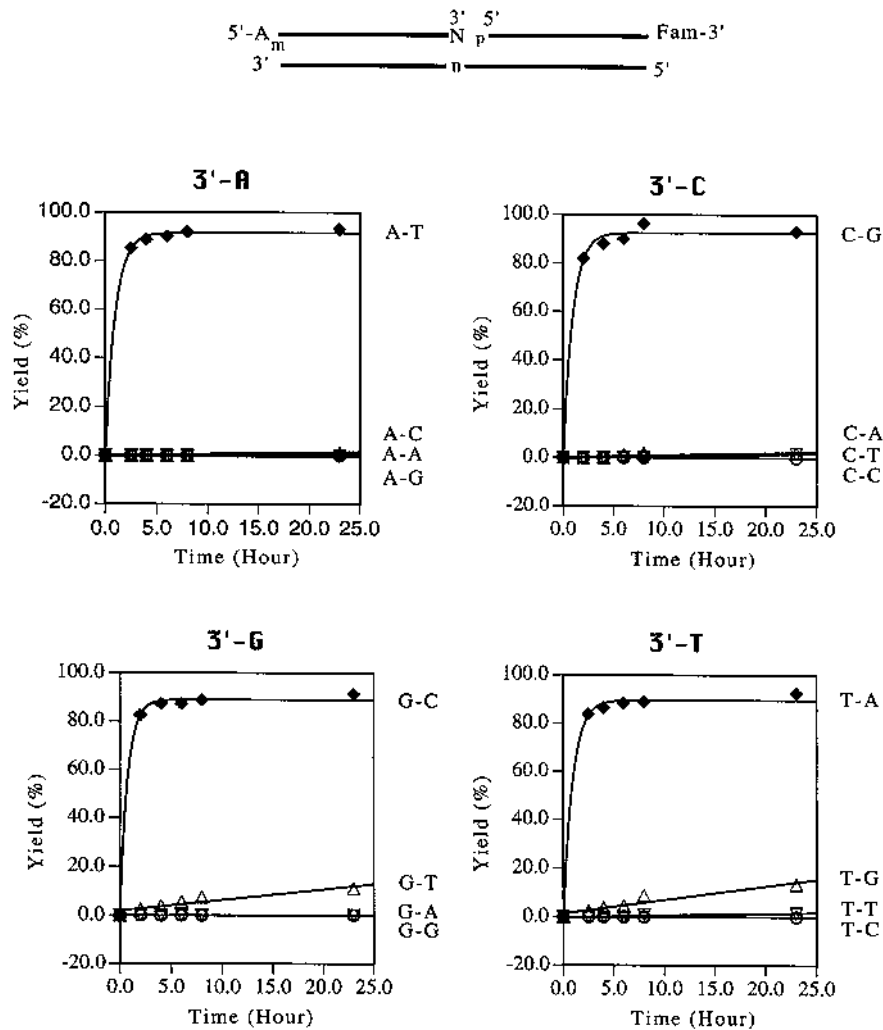


Figure 2. Fidelity of nick closure by *Tth* DNA ligase on the 3'-side of the nick. The ligase substrate (nicked DNA duplex), shown on the top of the figure, is formed by annealing the discriminating oligonucleotide LP3'(A, C, G or T) with a phosphorylated common oligonucleotide (3' fluorescently labeled, com3F) on the template strand. The discriminating base 'N' on the 3'-side of the discriminating oligonucleotide and the 'n' in the template strand were varied to give all 16 possible combinations of base pairing. A_m represents the 'A' tail at the 5'-end of the discriminating oligonucleotide. Reactions were carried out in a 40 ml mixture containing 1 mM NAD⁺, 12.5 nM (500 fmol) nicked DNA duplex substrates and 0.125 nM (5 fmol) *Tth* DNA ligase at 65°C. Aliquots (5 μl) were removed at 0, 2, 4, 6, 8 and 23 h and separated on denaturing polyacrylamide gels as described in Material and Methods. Data were analyzed using Genescan version 1.2 software. Results are plotted using Deltagraph Pro3 Software. Each panel represents results obtained with the same discriminating oligonucleotide, but with a different template strand. **3'-A)** The discriminating oligonucleotide was LP3'A. A-T (♦), A-C (Δ), A-A (∇) and A-G (○) represent DNA substrates containing TLg, CLg, ALg and GLg as the template strand respectively. **(3'-G)** The discriminating oligonucleotide was LP3'G. G-C (♦), G-T (Δ), G-A (∇) and G-G (○) represent DNA substrate with CLg, GLg, ALg and TLg as the template strand respectively. **(3'-C)** LP3'C was the discriminating oligonucleotide. C-G (♦), C-A (Δ), C-T (∇) and C-C (○) indicate DNA substrates containing GLg, ALg, TLg and CLg as the template strand respectively. **(3'-T)** LP3'T was the discriminating oligonucleotide. T-A (♦), T-G (Δ), T-T (∇) and T-C (○) represent DNA substrates containing ALg, GLg, Tlg and Clg as the template strand respectively.

single base mismatch destabilizing annealing of the octamer probe used on the 3'-side more than the tetradecamer probe used on the 5'-side, or (ii) an intrinsic feature of T4 DNA ligase (7). Our results support the second hypothesis, since the oligonucleotides used in our assay were similar in length and melting temperature. A detailed analysis of the vaccinia virus DNA ligase showed that all mismatches on the 5'-side were ligated more efficiently than mismatches on the 3'-side (12). Similar results have also been shown for DNA ligase from *S.cerevisiae*, although the mismatches tested on the 5'-side of the nick were not the same as those on the 3'-side of the nick (10). Also, the apurinic/aprimidinic (AP) sites on the 3'-side of nicks were less efficiently ligated by T4 DNA

ligase compared with the AP sites on the 5'-side of nicks (6). Therefore, the more stringent requirement for the canonical structure on the 3'-side of the nick compared with the 5'-side of the nick may be general to all DNA ligases.

The lower fidelity against T-G or G-T mismatches on the 3'-side by *Tth* and T4 DNA ligases mirrors the fidelity of DNA polymerases. The most common mispairs formed by insertion errors of DNA polymerase involve G pairing with T, although there were substantial variations observed depending on the DNA polymerase, method of assay and specific site investigated (5). It was also shown that G-T, A-C and G-A mispairings were the most frequent ones allowed by the *E.coli* polymerase III both *in vivo*

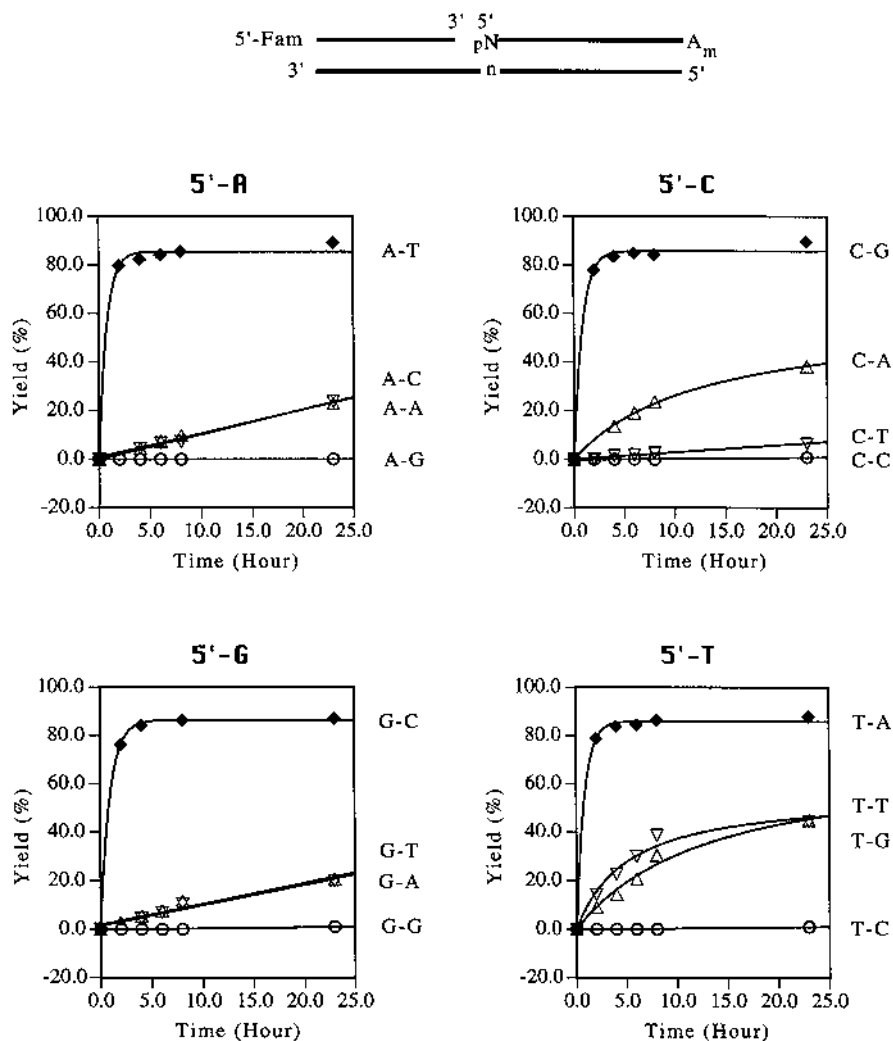


Figure 3. Fidelity of nick closure by a thermostable DNA ligase on the 5'-side of the nick. Reaction conditions were the same as in Figure 2 except that different discriminating and common oligonucleotides were used. The discriminating base was on the 5'-side of the nick (phosphorylated oligonucleotides RP5'A, C, G or T), while the common oligonucleotide was on the 3'-side of the nick and was 5'-labeled with FAM. (5'-A) The discriminating oligonucleotide was RP5'A. A-T (◆), A-C (Δ), A-A (∇) and A-G (○) represent DNA substrates containing TLg, CLg, ALg and GLg as the template strand respectively. (5'-G) The discriminating oligonucleotide was RP5'G. G-C (◆), G-T (Δ), G-A (∇) and G-G (○) represent DNA substrate with CLg, GLg, ALg and TLg as the template strand respectively. (5'-C) RP5'C was the discriminating oligonucleotide. C-G (◆), C-A (Δ), C-T (∇) and C-C (○) indicate DNA substrates containing GLg, ALg, TLg and CLg as the template strand respectively. (5'-T) RP5'T was the discriminating oligonucleotide. T-A (◆), T-G (Δ), T-T (∇) and T-C (○) represent DNA substrates containing ALg, GLg, TLg and Clg as the template strand respectively.

(33) and *in vitro* (34). Furthermore, *Taq* DNA polymerase extension of a 3' mismatched primer is more efficient for T-G and G-T mismatches than other mismatches at low dNTP concentrations (35). NMR and X-ray crystallography analysis of the structure of aberrant base pairs in duplex DNA oligonucleotide have indicated that G-T, A-C and G-A pairs existed as 'wobble' structures which differ slightly in dimensions from a normal Watson-Crick base pair (36,37). It was suggested that a well-stacked wobble pair may result in a more stable structure than a poorly stacked base pair approximating Watson-Crick geometry more closely. Both bacterial and mammalian DNA ligases have a very high fidelity against all purine-purine mismatches, including G-A or A-G mismatches on the 3'-side (9,11,12; this work), suggesting that base pair stability is not the predominant factor influencing ligase fidelity. Furthermore, the

related vaccinia virus DNA ligase and mammalian DNA ligase III both demonstrate even lower fidelity for C-T mismatches than G-T mismatches (11,12), while the *Tth* and T4 DNA ligases have a lower fidelity for G-T mismatches, with the *Tth* DNA ligase having virtually no misligations for C-T mismatches on the 3'-side (9; this work). Thus, the fidelity of DNA ligases is influenced not only by the compromise between increased stability and decreased helix distortion of a mismatched base pair, but also by specific structural determinants of the individual ligase protein.

Improving the fidelity of *Tth* DNA ligase

The fidelity of T4 DNA ligase has been expressed as a specificity ratio, defined as the ratio of ligation product formed in the

presence of matched versus mismatched template in the presence of 100 nM template at 30°C with 1 U enzyme (7). At high salt concentrations (200 mM) the specificity ratio increased from 6 to 60 for T-T mismatches and from 1.5 to 40 for A-A mismatches, although these more stringent conditions also increased the K_m 4-fold and decreased the V_{max} ~30-fold (7). The fidelity of *S.cerevisiae* DNA ligase I was defined as the ligation efficiency of substrates containing perfect matches versus mismatches in the presence of 3–900 ng enzyme at 20°C for 30 min (10). The efficiency of this DNA ligase was found to decrease 50-fold when a T-G mismatch was present on the 3'-side of the nick. The ligation efficiency of this ligase was also decreased >100-fold when a C-T mismatch was present of the 3'-side of the nick with a T-C mismatch on the 5'-side of the nick in the same substrate. It is difficult to compare the fidelity of different DNA ligases using either of the two terms described above, since they will vary when different incubation times and substrates are used.

In order to standardize comparisons with different substrates and enzymes, we define the fidelity of *Tth* DNA ligase as the ratio of the initial rate of perfect match ligation over the initial rate of mismatch ligation. Among substrates containing a mismatch on the 3'-side of the nick, T-G or G-T mismatches are the most difficult to discriminate and were therefore used as a standard assay for determining the fidelity ratio of *Tth* DNA ligase. Initial rates of C-G perfect match ligation over T-G mismatch ligation at 65°C gave a fidelity ratio of 4.5×10^2 for *Tth* DNA ligase (Table 1, line 1). Use of shorter discriminating probes ($T_m \sim 65^\circ\text{C}$) did not dramatically change perfect match ligations, but did decrease mismatch ligation. Thus, destabilizing the mismatched probe improved the fidelity ratio ~3-fold, to 1.5×10^3 (Table 1, line 2). This *Tth* DNA ligase fidelity ratio is at least 24-fold higher than reported for T4 ligase (7; on a T-T mismatch which is far easier to discriminate than the G-T mismatch used herein; see Fig. 2). One speculative explanation for the higher fidelity of *Tth* DNA ligase compared with mesophilic DNA ligases may be related to the lack of 3'→5' exonuclease (proofreading) activity in *Thermus* DNA polymerases (38). During DNA damage repair, *Thermus* DNA polymerase insertion of an incorrect base at the repair junction is infrequent, but when it does occur, the incorrect base cannot be removed by the polymerase. Consequently, the *Tth* DNA ligase may have evolved to have a higher stringency against sealing mismatches on the 3'-side of a nick, instead allowing an autonomous 3'→5' exonuclease to excise the erroneous base and maintain genome integrity.

This exquisite specificity of *Tth* DNA ligase has been exploited in LCR to detect a mutant allele (A-T match) in the presence of a 200-fold molar excess of wild-type sequence (G-T mismatch) (39). Several medical DNA detection problems will require an even greater specificity. We have devised two approaches to further improve the fidelity ratio of *Tth* DNA ligase by destabilizing the enzyme-substrate interactions.

Improving the fidelity of *Tth* DNA ligase by modifying the DNA substrate

An elegant method for improving allele-specific PCR is based on using primers with a deliberate mismatch adjacent to the discriminating 3' base (40,41). This destabilizing mismatch did not dramatically reduce *Taq* polymerase extension of the correct target allele, but owing to a double mismatch of the other allele,

the extension efficiency of the incorrect allele was reduced by a factor of 100- to 1000-fold (40).

We adapted this approach for use in ligation reactions by deliberately introducing an A-C mismatch at the *third position* on the 3'-side of the nick (shown in bold in both perfect match and mismatch DNA substrates; Table 1, line 3). The original perfect match substrate now has a single mismatch and the original mismatch substrate now has two mismatches (one right on the 3'-side of the nick, the other three bases in on the 3'-side of the same probe). This A-C mismatch in the third position reduced the ligation efficiency of a 3' matched (C-G) substrate almost 10-fold with 5 fmol *Tth* DNA ligase (data not shown). In order to obtain an initial rate comparable with that with normal perfect match ligations, the amount of enzyme was increased to 50 fmol. The ligase fidelity ratio increased by ~4-fold (to 5.8×10^3) when the extra mismatch was introduced (Table 1, line 3). The internal mismatch has a greater destabilizing effect on the structure near the nick for the probe containing the 3' mismatch on the nick than the probe containing the perfect match on the nick. Therefore, the overall fidelity of the ligase was improved. Similar results (4-fold increase) were also obtained when a T-G or G-T mismatch was introduced at the same third position on the 3'-side of the nick (data not shown). If the extra mismatch (C-A) was introduced into the second position on the 3'-side of the nick, ligation of the perfect match on the 3'-side of the nick by *Tth* DNA ligase was strongly inhibited (175-fold lower than with no mismatch at the second position). In contrast, *Taq* DNA polymerase was not affected by the extra mismatch at the second position from the 3'-end of the primer (40). A key difference between PCR and LCR is that an adjacent mismatch only affects extension from the target during the initial PCR cycle, but affects ligation on the target and products during every cycle of the LCR reaction.

As an alternative to a mismatched base pair we hypothesized that a universal nucleotide analog might maintain DNA helix integrity with a perfect match substrate while still destabilizing a nearby mismatch and thus further improve ligation fidelity. Oligonucleotides containing the nucleotide analog 3-nitropyrrole deoxyribonucleotide (**Q**) at multiple sites in place of the natural nucleotides have been shown to function effectively as sequencing and PCR primers (28,42). 3-Nitropyrrole presumably allows preservation of helix integrity because it is sufficiently small to fit opposite any of the four natural bases and has high stacking potential due to a highly polarized π electronic configuration. T_m studies on **Q**-containing oligonucleotides indicate that **Q** base pairs with relatively little discrimination (ΔT_m 3°C), but the stability of **Q**-A (most stable), **Q**-T, **Q**-C and **Q**-G (least stable) base pairs is significantly less than that of an A-T or C-G base pair. Consequently, **Q**, if located 3 nt in from the 3'-end of a probe, could significantly enhance local melting if it were present in conjunction with a mismatch at the 3'-position, while at the same time preserving helix integrity more than a mismatch when present in conjunction with a base pair match at the 3'-end. When sequences SLP3'**Q**TC and SLP3'**Q**TT were tested as ligation substrates for *Tth* DNA ligase, an even better fidelity ratio (9-fold increase) was obtained (Table 1, line 4). Note that the base on the template strand was changed from C to A. As determined by the initial rates of perfect match ligation, the **Q** base analog appears to pair with A and T equally well, less well with C and very poorly with G. The ratio of initial rates obtained when **Q** was paired with different bases is T:A:C:G = 23:16:5:1. As observed with C/A mismatches above, when the **Q** analog is located at the second

Table 1. Improving the fidelity of *Tth* DNA ligase

Discriminating Probes ^a	Ligase (conc.)	Ligation Substrate ^b	Initial Rates ^c C-G match (fmol/min)	Initial Rates ^c T-G mismatch (fmol/min)	Ligation ^d fidelity $\frac{\text{Rate}_{\text{match}}}{\text{Rate}_{\text{mismatch}}}$
Long Probes (LP3'C and LP3'T)	Wild Type (0.25 nM)	$\begin{array}{l} \text{---GTN p---F} \\ \text{---CAG} \end{array}$	25	5.5×10^{-2}	4.5×10^2
Short Probes (SLP3'C and SLP3'T)	Wild Type (0.25 nM)	$\begin{array}{l} \text{---GTN p---F} \\ \text{---CAG} \end{array}$	29	2.0×10^{-2}	1.5×10^3
Probes with deliberate mismatches (SLP3'ATC and SLP3'ATT)	Wild Type (2.5 nM)	$\begin{array}{l} \text{---ATN p---F} \\ \text{---CAG} \end{array}$	26	4.4×10^{-3}	5.9×10^3
Probes with Q base analogue (SLP3'QTC and SLP3'QTT)	Wild Type (2.5 nM)	$\begin{array}{l} \text{---QTN p---F} \\ \text{---AAG} \end{array}$	44	3.3×10^{-3}	1.3×10^4
Short Probes (SLP3'C and SLP3'T)	K294R (2 nM)	$\begin{array}{l} \text{---GTN p---F} \\ \text{---CAG} \end{array}$	65	1.2×10^{-2}	5.4×10^3
Short Probes (SLP3'C and SLP3'T)	K294P (4 nM)	$\begin{array}{l} \text{---GTN p---F} \\ \text{---CAG} \end{array}$	50	3.1×10^{-3}	1.6×10^4
Probes with Q base analogue (SLP3'QTC and SLP3'QTT)	K294R (30 nM)	$\begin{array}{l} \text{---QTN p---F} \\ \text{---AAG} \end{array}$	51	1.8×10^{-3}	2.8×10^4
Probes with Q base analogue (SLP3'QTC and SLP3'QTT)	K294P (120 nM)	$\begin{array}{l} \text{---QTN p---F} \\ \text{---AAG} \end{array}$	38	9.0×10^{-4}	4.2×10^4

Reactions were carried out as described in Materials and Methods. The concentration (conc.) of *Tth* DNA ligase used in each experiment is indicated. In assays with wild-type enzyme, results were calculated as the average of two or more experiments with a standard deviation of <15%.

^aSequences of probes LP3'C, LP3'T and Com 3F are shown in Figure 1. Sequences for other discriminating probes are as the follows: SLP3'C, 5'-TACGTCTGCGGTGTTGCGTC-3'; SLP3'T, 5'-CGTCTGCGGTGTTGCGTT-3'; SLP3'ATC, 5'-ATGCGTCTGCGGTGTTGCATC-3'; SLP3'ATT, 5'-GCGTCTGCGGTGTTGCATT-3'; SLP3'QTC, 5'-AAATGCGTCTGCGGTGTTGCQTC-3'; SLP3'QTT, 5'-ATGCGTCTGCGGTGTTGCQTT-3'.

^bBase 'N' in the discriminating oligonucleotide represents either C or T. Q indicates the Q base analog. The template strand for all substrates tested except those containing Q base analogs is GLg and its sequence is shown in Figure 1. The template strand in substrates containing Q base analogs is GLg.m3A, which differs from GLg at a single site shown in bold.

^cInitial rates of ligation (fmol/min) were calculated as the slope of the linear graph with the x-axis as the time in min and the y-axis as the amount of products in fmol.

^dLigation fidelity of *Tth* DNA ligase is defined as the ratio of the initial rate of perfect match ligation over the initial rate of mismatch ligation.

position, the ligation rate is also strongly inhibited (55-fold lower than with no analog at the second position).

On the basis of modeling studies (QUANTA/CHARMM), Q can fit opposite both T and A with minimal perturbation of the helix structure (28). In one case Q would assume the *anti* conformation (Q-T) and in the other the *syn* conformation (Q-A). Nevertheless, studies in progress suggest that hydrogen bonding plays only a minor role in 3-nitropyrrole-natural base interactions (Klewer and Bergstrom, unpublished).

Improving the fidelity of *Tth* DNA ligase by site-directed mutagenesis of the ligase protein

The fidelity of DNA polymerases may be decreased by site-directed mutagenesis of motifs associated with primer-template binding (HIV polymerase) (43) or the exoIII motif of ϕ 29 DNA

polymerase (44), T4 DNA polymerase (45,46) or human DNA polymerase α (47). Occasionally, this same exoIII motif or motif 'A' yields *increased* fidelity mutants, also known as 'anti-mutator' strains, which reflects the complex interplay between the polymerizing and 3'→5' exonuclease activities of these enzymes in modulating overall fidelity (45,46,48,49).

Mutant *Tth* DNA ligases (32) which retained near wild-type nick closing activity were assayed for changes in fidelity. Two mutant ligases, K294R and K294P, were shown to have increased their fidelity ratios (Table 1, lines 5–8). With regular substrates, the fidelity increased ~4-fold with K294R and 11-fold with K294P. When Q base analogs were used together with mutant *Tth* DNA ligases, the fidelity of the ligation reactions were increased by ~20-fold (K294R and K294P), although higher concentrations of mutant ligase were required.

In summary, we have developed a quantitative fluorescence assay for analyzing the fidelity of *Tth* DNA ligase. This enzyme exhibits significantly greater discrimination against all single base mismatches on the 3'-side of the nick in comparison with those on the 5'-side of the nick. Among all 12 possible single base pair mismatches on the 3'-side of the nick, only T-G and G-T mismatches generated a quantifiable level of ligation products after extended incubation. The fidelity of *Tth* DNA ligase can be improved further by designing discriminating oligonucleotides with melting temperature values near the ligation temperature of 65°C, by introducing a deliberate mismatch base or a nucleotide analog into the left third position on the 3'-side of the nick and/or by using mutant *Tth* DNA ligase.

Numerous medical problems will require exquisite single base discrimination, for example finding rare cancer cells among many normal cells, distinguishing pathogenic microorganisms among normal flora, identifying oncogenic HPV strains in mixed infections and detecting the emergence of drug-resistant organisms. While our model studies were limited to either pure mismatched or matched target substrates, we have recently used nucleotide analogs and mutant *Tth* DNA ligases to improve discrimination of a minor correct target in a mixture of incorrect target, which more closely mimics DNA detection in biological samples (M.Khanna, J.Day, P.Paty, D.Bergstrom and F.Barany, unpublished results).

ACKNOWLEDGEMENTS

The authors gratefully acknowledge Drs Stewart Shuman, George Barany, Robert Hammer, Pat Paty, Emily Winn-Deen, Alan Mayer, Marilyn Khanna, Joe Day, Herman Block and Monib Zirvi for many thoughtful discussions. We especially thank Drs Weiguo Cao and Darren Day for many valuable ideas and suggestions. We thank Peiming Zhang for synthesis of 3-nitropyrrole deoxyribonucleoside and its phosphoramidite. We thank Ms Jing Lu and Ms Antje Koller for their excellent technical assistance. FB is supported in part by a Hirsch/Monique Weill-Caulier Career Scientist Award. This work was supported by grants from the Applied Biosystems Division, Perkin Elmer Corporation (FB), the Strang Cancer Prevention Center (FB) and the National Institutes of Health (GM-41337-06 to FB and AI-366-01 to DB).

REFERENCES

- Lehman, I.R. (1974) *Science*, **186**, 790–797.
- Barnes, D.E., Tomkinson, A.E., Lehmann, A.R., Webster, A.D.B. and Lindahl, T. (1992) *Cell*, **69**, 495–503.
- Loeb, L.A. and Kunkel, T.A. (1982) *Annu. Rev. Biochem.*, **52**, 429–457.
- Kunkel, T.A. and Bebenek, K. (1988) *Biochim. Biophys. Acta*, **95**, 1–15.
- Echols, H. and Goodman, M.F. (1991) *Annu. Rev. Biochem.*, **60**, 477–511.
- Goffin, C., Bailly, V. and Verly, W.G. (1987) *Nucleic Acids Res.*, **15**, 8755–8771.
- Wu, D.Y. and Wallace, R.B. (1989) *Gene*, **76**, 245–254.
- Harada, K. and Orgel, L. (1993) *Nucleic Acids Res.*, **21**, 2287–2291.
- Landegren, U., Kaiser, R., Sanders, J. and Hood, L. (1988) *Science*, **241**, 1077–1080.
- Tomkinson, A.E., Tappe, N.J. and Friedberg, E.C. (1992) *Biochemistry*, **31**, 11762–11771.
- Husain, I., Tomkinson, A.E., Burkhart, W.A., Moyer, M.B., Ramos, W., Mackey, Z., Besterman, J.M. and Chen, J. (1995) *J. Biol. Chem.*, **270**, 9683–9690.
- Shuman, S. (1995) *Biochemistry*, **34**, 16138–16147.
- Barany, F. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 189–193.
- Day, D., Speiser, P.W., White, P.C. and Barany, F. (1995) *Genomics*, **29**, 152–162.
- Eggerding, F.A. (1995) *PCR Methods Applicat.*, **4**, 337–345.
- Eggerding, F.A., Iovannisci, D.M., Brinson, E., Grossman, P. and Winn-Deen, E.S. (1995) *Hum. Mutat.*, **5**, 153–165.
- Feero, W.T., Wang, J., Barany, F., Zhou, J., Todorovic, S.M., Conwit, R., Galloway, G., Hausmanowa-Petrusewicz, I., Fidzianska, A., Arahata, K., Wessel, H.B., Wadelius, C., Marks, H.G., Hartlage, P., Hayakawa, H. and Hoffman, E.P. (1993) *Neurology*, **43**, 668–673.
- Frenkel, L.M., Wagner, L.E., Atwood, S.M., Cummins, T.J. and Dewhurst, S. (1995) *J. Clin. Microbiol.*, **33**, 342–347.
- Grossman, P.D., Bloch, W., Brinson, E., Chang, C.C., Eggerding, F.A., Fung, S., Iovannisci, D.A., Woo, S. and Winn-Deen, E.S. (1994) *Nucleic Acids Res.*, **22**, 4527–4534.
- Iovannisci, D.M. and Winn-Deen, E.S. (1993) *Mol. Cell. Probes*, **7**, 35–43.
- Prchal, J.T., Guan, Y.L., Prchal, J.F. and Barany, F. (1993) *Blood*, **81**, 269–271.
- Ruiz-Opaz, N., Barany, F., Hirayama, K. and Herrera, V. (1994) *Hypertension*, **24**, 260–270.
- Wiedmann, M., Czajka, J., Barany, F. and Batt, C. (1992) *Appl. Environ. Microbiol.*, **58**, 3443–3447.
- Wiedmann, M., Barany, F. and Batt, C.A. (1993) *Appl. Environ. Microbiol.*, **59**, 2743–2745.
- Winn-Deen, E., Grossman, P., Fung, S., Woo, S., Chang, C., Brinson, E. and Eggerding, F. (1993) *Am. J. Hum. Genet.*, **53**, 1512.
- Winn-Deen, E.S., Adriano, T., Bloch, W., Brinson, E., Chang, C., Fung, S., Iovannisci, D.M., Madonik, A., Sherman, D.G. and Woo, S. (1994) *Clin. Chem.*, **40**, 1092.
- Zebala, J. and Barany, F. (1995) In Innis, M.A., Gelfand, D.H. and Sninsky, J.J. (eds), *PCR Strategies*. Academic Press, San Diego, CA, pp. 335–346.
- Bergstrom, D.E., Zhang, P., Toma, P.H., Andrews, P.C. and Nichols, R. (1995) *J. Am. Cancer Soc.*, **117**, 1201–1209.
- Barany, F. and Gelfand, D. (1991) *Gene*, **109**, 1–11.
- Takahashi, M., Yamaguchi, E. and Uchida, T. (1986) *Agric. Biol. Chem.*, **50**, 1333–1334.
- Bradford, M.M. (1976) *Anal. Biochem.*, **72**, 248–254.
- Luo, J. and Barany, F. (1996) *Nucleic Acids Res.*, **24**, 3079–3085.
- Schaaper, R.M. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 8126–8130.
- Sloane, D.L., Goodman, M.F. and Echols, H. (1988) *Nucleic Acids Res.*, **16**, 6465–6475.
- Kwok, S., Kellogg, D.E., Spasic, D., Goda, L., Levenson, C. and Sninsky, J.J. (1990) *Nucleic Acids Res.*, **18**, 999–1005.
- Hunter, W.N., Brown, T., Anand, N.N. and Kennard, O. (1986) *Nature*, **320**, 552–555.
- Patel, D.J., Kozlowski, S.A., Ikuta, S. and Itakura, K. (1984) *Fedn. Proc. Fedn. Am. Soc. Exp. Biol.*, **43**, 2663–2670.
- Lawyer, F.C., Stoffel, S., Saiki, R.K., Myambo, K., Drummond, R. and Gelfand, D.H. (1989) *J. Biol. Chem.*, **264**, 6427–6437.
- Balles, J. and Pflugfelder, G.O. (1994) *Mol. Gen. Genet.*, **245**, 734–740.
- Cha, R.S., Zarbl, H., Keohavong, P. and Thilly, W.G. (1992) *PCR Methods Applicat.*, **2**, 14–20.
- Rust, S., Funke, H. and Assman, G. (1993) *Nucleic Acids Res.*, **21**, 3623–3629.
- Nichols, R., Andrews, P.C., Zhang, P. and Bergstrom, D.E. (1994) *Nature*, **369**, 492–493.
- Beard, W.A., Stahl, S.J., Kim, H.-R., Bebenek, K., Kumar, A., Strub, M.-P., Becerra, S.P., Kumkel, T.A. and Wilson, S.H. (1994) *J. Biol. Chem.*, **269**, 28091–28097.
- Soengas, M.S., Esteban, J.A., Lazaro, J.M., Bernad, A., Blasco, M.A., Salas, M. and Blanco, L. (1992) *EMBO J.*, **11**, 4227–4237.
- Reha-Krantz, L.J., Nonay, R.J. and Stocki, S. (1993) *J. Virol.*, **67**, 60–66.
- Reha-Krantz, L.J. and Nonay, R.L. (1994) *J. Biol. Chem.*, **269**, 5635–5643.
- Dong, Q., Copeland, W.C. and Wang, T.S.-F. (1993) *J. Biol. Chem.*, **268**, 24163–24174.
- Dong, Q., Copeland, W.C. and Wang, T.S.-F. (1993) *J. Biol. Chem.*, **268**, 24175–24182.
- Copeland, W.C., Lam, N.K. and Wang, T.S.-F. (1993) *J. Biol. Chem.*, **268**, 11041–11049.
- Rychlik, W., Domier, L.L., Gardner, P.R., Hellmann, G.M. and Rhoads, R.E. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 945–949.
- Breslauer, K.J., Frank, R., Blocker, H. and Markey, L.A. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 3746–3750.