

# Biochemical properties of a high fidelity DNA ligase from *Thermus* species AK16D

Jie Tong, Weiguo Cao and Francis Barany\*

Department of Microbiology, Hearst Microbiology Research Center, Strang Cancer Prevention Center, The Joan and Sanford I. Weill Medical College of Cornell University, 1300 York Avenue, Box 62, New York, NY 10021, USA

Received September 29, 1998; Revised and Accepted November 28, 1998

DDBJ/EMBL/GenBank accession nos AF092862–AF092868

## ABSTRACT

**NAD<sup>+</sup>-dependent DNA ligases from thermophilic bacteria *Thermus* species are highly homologous with amino acid sequence identities ranging from 85 to 98%. *Thermus* species AK16D ligase, the most divergent of the seven *Thermus* isolates collected worldwide, was cloned, expressed in *Escherichia coli* and purified to homogeneity. This *Thermus* ligase is similar to *Thermus thermophilus* HB8 ligase with respect to pH, salt, NAD<sup>+</sup>, divalent cation profiles and steady-state kinetics. However, the former is more discriminative toward T/G mismatches at the 3'-side of the ligation junction, as judged by the ratios of initial ligation rates of matched and mismatched substrates. The two wild-type *Thermus* ligases and a *Tth* ligase mutant (K294R) demonstrate 1–2 orders of magnitude higher fidelity than viral T4 DNA ligase. Both *Thermus* ligases are active with either the metal cofactor Mg<sup>2+</sup>, Mn<sup>2+</sup> or Ca<sup>2+</sup> but not with Co<sup>2+</sup>, Ni<sup>2+</sup>, Cu<sup>2+</sup> or Zn<sup>2+</sup>. While the nick closure step with Ca<sup>2+</sup> becomes rate-limiting which results in the accumulation of DNA–adenylate intermediate, Ni<sup>2+</sup> only supports intermediate formation to a limited extent. Both *Thermus* ligases exhibit enhanced mismatch ligation when Mn<sup>2+</sup> is substituted for Mg<sup>2+</sup>, but the *Tsp*. AK16D ligase remains more specific toward perfectly matched substrate.**

## INTRODUCTION

DNA ligases, as an essential component of DNA replication, recombination and repair systems found from viruses to humans, catalyze the formation of a phosphodiester bond at single-stranded breaks on duplex DNA (1). DNA ligases can be classified into two families based on cofactor dependence. ATP-dependent ligases are found in bacteriophages (2,3), *Chlorella* virus PBCV-1 (4), vaccinia virus (5), Archea (6,7), yeasts (8–10), mammals (11–13) and more recently eubacteria (14,15). NAD<sup>+</sup>-dependent ligases, however, are found exclusively in eubacteria. While some higher eukaryotic organisms may use multiple ATP-dependent ligases to fulfil diverse biological functions, some simple eubacteria genomes appear to encode

both NAD<sup>+</sup>- and ATP-dependent ligases (15,16). The origin of the additional ATP-dependent ligases in these genomes remains to be determined.

Although the ATP-dependent ligases and NAD<sup>+</sup>-dependent ligases share little sequence homology, all the ligases investigated so far use the same KXDG motif to form adenylated enzyme intermediate (11,17,18) and, furthermore, they seem to be organized by a similar structural fold in the core of the ATP/NAD<sup>+</sup>-binding domain (19–22). The diversity of ligase sequences is not only reflected by their different optimal reaction conditions and kinetic rates, but more importantly by their different specificities toward matched and mismatched substrates. Among the viral ATP-dependent ligases, the broad substrate tolerance is represented by the T4 enzyme which seals various mismatches on both the 3'- and 5'-side of the nick junction (23). Vaccinia ligase ligates various mismatches at both the 3'-hydroxyl and 5'-phosphate sides with the exception of purine-purine mismatch pairs at the 3'-hydroxyl side (5). Mammalian ATP-dependent ligases show different substrate sensitivity, as ligase I is more sensitive to 3' mismatches than ligase III (24). Additionally, both ligase I and III tolerate a 3'-C/T mismatch more than a 3'-G/T mismatch. Little is known about archeal ATP-dependent ligases which may be the progenitor of eukaryotic ATP-dependent ligases. Studies on NAD<sup>+</sup>-dependent DNA ligase from *Escherichia coli*, along with T4 ligase, have contributed immensely to our understanding of the basic biochemical pathway of DNA ligation reaction (1,25). Studies on the NAD<sup>+</sup>-dependent ligase from *Thermus thermophilus* HB8 has revealed the high discriminative power this enzyme possesses (26). Although mismatches at the 5'-phosphate side are tolerated to some degree (5'-A/C, 5'-A/A, 5'-C/A, 5'-C/T, 5'-G/T, 5'-G/A, 5'-T/T and 5'-T/G), mismatches at the 3'-hydroxyl side essentially abolish nick-closure activity except 3'-G/T and 3'-T/G mismatches (26). Apparently, sequence divergence and subsequent subtle structural variation among DNA ligases underlie an enzyme's recognition preferences toward different mismatch base pairs.

The study of ligase biochemistry is not only important for understanding its biological functions, but also for developing new technologies. The single nucleotide discrimination observed on DNA ligases has led to the development of ligase-mediated detection techniques (23,27–31). Ligase-based linear signal amplification (Ligase Detection Reaction, LDR), combined with

\*To whom correspondence should be addressed. Tel: +1 212 746 6509; Fax: +1 212 746 8587; Email: barany@mail.med.cornell.edu

PCR-based gene-specific target amplification, has been proven to be a powerful tool in cancer and disease gene mutation detection (32,33). The PCR/LDR technique relies on two properties of a DNA ligase: specificity and thermostability. *Tth* DNA ligase has been successfully used in LDR and LCR due to its highly discriminative nick closure activity toward a perfect matched substrate and its thermostability which makes thermocycling possible (30,31). To date, one more ligase has been cloned and sequenced from *Thermus scotoductus* (34,35), but the substrate specificity of this ligase was not determined.

The sensitivity of LDR is dependent on the fidelity of the ligase. This work intends to explore the natural ligases from *Thermus* species collected worldwide and evaluate their sequence divergence and biochemical properties. We herein report the sequence analysis of seven *Thermus* DNA ligase genes, cloning and expression of a ligase from *Thermus* species AK16D and biochemical characterization of this high fidelity enzyme.

## MATERIALS AND METHODS

### Reagents, media, strains and GenBank accession nos

All routine chemical reagents were purchased from Sigma Chemical Corp. (St Louis, MO) or Fisher Scientific (Fair Lawn, NJ). Restriction enzymes and T4 DNA ligase were obtained from New England Biolabs (Beverly, MA). Oligonucleotide synthesis reagents, DNA sequencing kits and PCR kits were obtained from Applied Biosystems Division of Perkin Elmer Corp. (Foster City, CA). dNTPs, BSA and ATP were purchased from Boehringer-Mannheim (Indianapolis, IN). *Pfu* DNA polymerase was purchased from Stratagene (La Jolla, CA). *Escherichia coli* strain NovaBlue(DE3)pLysS and plasmid pET11c were purchased from Novagen Inc. (Madison, WI). The protein assay kit was from Bio-Rad (Hercules, CA). The HiTrap Blue affinity column was from Pharmacia (Piscataway, NJ). LB medium was prepared according to the standard formula (36). Sonication buffer consisted of 50 mM Tris-HCl, pH 8.0, and 1 mM EDTA. TE buffer consisted of 10 mM Tris-HCl, pH 8.0, and 1 mM EDTA. Stop solution consisted of 50 mM EDTA, 80% formamide and 1% Blue Dextran. *Tth* DNA ligase and its mutant K294R were purified as previously described (26). GenBank accession nos for *Thermus* ligases: *Tsp.* AK16D ligase gene, AF092862; *Thermus aquaticus* YT-1, AF092863; *Tth* HB8, 118780; *Thermus flavus*, AF092864; *Thermus filiformis* Tok4A2, AF092865; *T.filiformis* Tok6A1, AF092866; *Tsp.* SM32, AF092868; *Tsp.* Vil3, AF092867.

### Oligodeoxyribonucleotide synthesis

Oligonucleotides were synthesized by using a 394 automated DNA synthesizer from Applied Biosystems Division of Perkin Elmer Corp. PCR and sequencing primers were purified by ethanol precipitation according to the instruction manual. The degenerate sense primer d[ATC(T/A)(C/G)CGACGC(C/G)GA-(G/A)TA(T/C)GA] corresponding to amino acids 32–38 (ISDAEYD) in the *T.thermophilus* HB8 DNA ligase gene and antisense primers d[CC(C/G)GT(C/G)C(G/T)(G/C)CC(G/C)AC-(C/T)TG(A/G)AA] and d[GCCTTCTC(C/G/A)A(A/G)(T/C)-TTG(C/G)(A/T)(G/C)CC] corresponding to amino acids 333–339 (FQVGRGTG) and 641–647 (GSKLEKA) were used to amplify DNA ligase gene fragments from *Thermus* strains. Additional PCR and sequencing primers were synthesized as required. The

sequences of PCR amplification primers for cloning the *Tsp.* AK16D DNA ligase gene into pET11c vector were d(GCGAT-TTCATATGACCCTAGAGGAGGCCCG) and d(GCGGGA-TCCGAGGCCTTGGAGAAGCTCTT), where the *Nde*I and *Bam*HI sites are underlined and the initiation codon in the forward primer is shown in bold. Oligonucleotide substrates for ligation assay were purified on a denaturing sequencing gel (7 M urea/10% polyacrylamide) (37). Phosphorylation at the 5'-terminus of oligonucleotides was achieved during synthesis by using Chemical Phosphorylation Reagent (Glen Research, Sterling, VA). A fluorescent group was attached to the 3'-terminus using a fluorescein CPG column (Glen Research).

### DNA amplification, cloning and sequence analysis

Genomic DNAs from *Thermus* strains were isolated as previously described (38). PCR amplifications with degenerate and unique primers and inverse PCR on circularized templates were carried out in a GeneAmp PCR System 9700 thermocycler (Applied Biosystems Division of Perkin Elmer) as described (39). The nucleotide sequences of amplified ligase fragments were directly determined on an ABI 373 sequencer using an ABI PRISM™ Dye Terminator Cycle Sequencing Ready Reaction Kit (Perkin Elmer). Full-length *Tsp.* AK16D DNA ligase gene was amplified using PCR amplification primers as described above, digested with *Nde*I and *Bam*HI, ligated into the cloning vector pET11c treated with the same pair of restriction enzymes and transformed into *E.coli* strain NovaBlue(DE3)pLysS. Inserts in pET expression vectors were sequenced in both orientations to ensure that the plasmid constructs were free of PCR or ligation error. Nucleic acid and protein sequence analyses were carried out by the Clustal method (40) using the MegAlign program of DNASTAR (Madison, WI).

### Expression and purification of *Tsp.* AK16D DNA ligase

*Escherichia coli* NovaBlue(DE3)pLysS cells containing plasmid pTAK encoding the *Tsp.* AK16D DNA ligase gene from a pET11c construct was propagated overnight at 37°C in LB medium containing 50 µg/ml ampicillin, 25 µg/ml chloramphenicol and 0.2% glucose. Overnight cultures were diluted 100-fold into the same medium, grown until the optical density of the culture reached 0.5 at 600 nm, then induced by the addition of IPTG to a final concentration of 1 mM and grown for an additional 4 h under the same conditions. Cells were collected by centrifugation, frozen/thawed at –20/23°C, disrupted by sonication and clarified by centrifugation as previously described (39). The resulting supernatants were heated at 70°C for 15 min to denature thermolabile *E.coli* proteins, placed on ice for 30 min to aggregate the denatured proteins and cleared of denatured proteins by microcentrifugation for 15 min at 4°C. The partially pure DNA ligase was further purified by chromatography using 1 ml HiTrap Blue affinity column. Briefly, the column containing *Tsp.* AK16D DNA ligase was washed extensively with TE buffer (pH 7.8) containing 0.1 M NaOAc and the ligase was eluted with TE buffer (pH 7.8) containing 2 M NaCl. After dialysis against TE buffer (pH 8.0) containing 0.2 M KCl and concentration using Centricon-30 (Amicon), protein concentration was assayed by the Bradford method with reagents supplied in the Bio-Rad protein assay kit. The amount of protein was determined using BSA as the standard. Ligase purity was verified through 7.5%

SDS-PAGE analysis followed by visualizing the overloaded gel with routine Coomassie Brilliant Blue R staining.

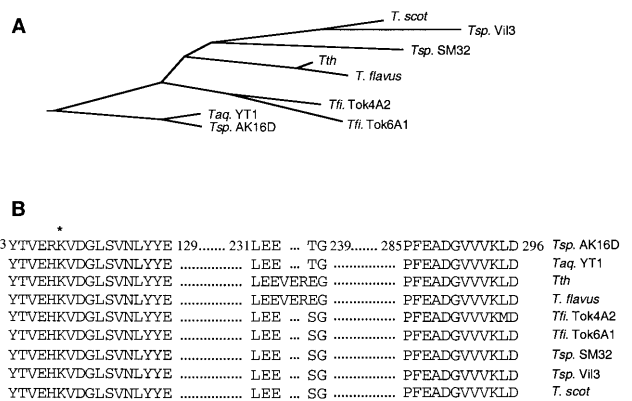
### Substrates and ligation assay

The oligonucleotide perfect matched substrate was formed by annealing two short oligonucleotides (33mer for LP3'C and 30mer for Com3F) with a 59mer complementary oligonucleotide (Glg). Oligonucleotides LP3'C and Glg were in 1.5-fold excess so that all the 3'-Fam-labeled Com3F represented nicked substrates (details in 26). The T/G mismatched substrate was formed by annealing LP3'T, which introduced a single base pair mismatch at the 3'-end of the nick junction, along with Com3'F to the complementary strand (Glg). The nicked DNA duplex substrates were formed by denaturing DNA probes at 94°C for 2 min followed by re-annealing at 65°C for 2 min in ligation buffer. The sequences of the oligonucleotides are listed in Scheme 1 (p represents the 5'-phosphate group).

Ligation mixtures (20  $\mu$ l) containing the indicated amount of DNA ligase and matched or mismatched substrate in the ligation buffer (20 mM Tris-HCl, pH 7.6 at room temperature, 10 mM MgCl<sub>2</sub>, 100 mM KCl, 10 mM dithiothreitol, 1 mM NAD<sup>+</sup> and 20  $\mu$ g/ml BSA) were incubated at 65°C for a predetermined time. Reactions were terminated by the addition of an equal volume of stop solution. Samples (5  $\mu$ l) were separated by electrophoresis through an 8 M urea-10% polyacrylamide GeneScan gel according to the manufacturer's instruction manual (Applied Biosystems Division of Perkin Elmer). The unreacted substrates (30mer Com3F) and products (63mer) were quantified using GeneScan analysis software 672 v.2.0 (Applied Biosystems Division of Perkin Elmer). For initial rate measurement, the reaction mixture consisted of 12.5 nM nicked DNA duplex substrates and the indicated amount of DNA ligases in ligation reaction buffer. Aliquots of 5  $\mu$ l from a 160  $\mu$ l reaction mixture were removed at 0, 10, 20, 30, 40, 50 and 60 s for reactions containing matched substrates and at 0, 1, 2, 3, 4, 5 and 6 h for reactions containing mismatched substrates and mixed with 5  $\mu$ l of stop solution. The results were plotted using DeltaGraph Pro3 software (DeltaPoint Inc., Monterey, CA). The initial rates were determined as the slope of linear range of the graph with the *x*-axis as the time and the *y*-axis as the amount of the ligation product generated.

### Steady-state kinetics

Steady-state kinetic constants were determined by measuring initial rates of the ligation reaction at a given substrate concentration (nicked DNA duplex substrate concentration ranging from 25 to 400 nM) and a given ligase concentration (12.5 pM for both *Tth* and *Tsp*. AK16D) in 100  $\mu$ l reaction volume at 65°C. A 5  $\mu$ l aliquot was removed at 0, 2, 4, 6, 8 and 10 min and mixed with 5  $\mu$ l of stop solution. The remaining substrate was separated from ligated product by GeneScan gel as described above. Initial rates of the ligation reactions were calculated from the generation of ligated product over time. The  $K_m$  and  $k_{cat}$  values were determined using Ultrafit (Biosoft, Ferguson, MO).

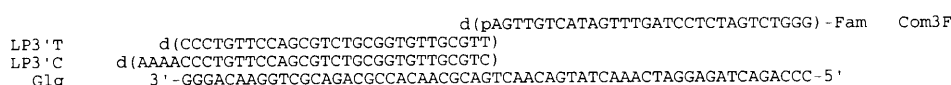


**Figure 1.** Sequence comparison of *Thermus* DNA ligases. (A) Evolutionary tree for *Thermus* DNA ligases. (B) Regional sequence alignment of nine *Thermus* ligases. The amino acid sequence of *T.scotoductus* is retrieved from GenBank by accession no. 1085749. The adenylation motif KXDG is underlined and the adenylation site is marked by \*. The numbering of amino acids is based on *Tsp*. AK16D ligase. The complete sequence of *Tsp*. AK16D ligase and partial sequences of six other *Thermus* ligases have been deposited with GenBank under accession nos AF092862 for *Tsp*. AK16D ligase gene, AF092863 for *T.aquaticus* YT-1, 118780 for *Tth* HB8, AF092864 for *T.flavus*, AF092865 for *T.filiformis* Tok4A2, AF092866 for *T.filiformis* Tok6A1, AF092868 for *Tsp*. SM32 and AF092867 for *Tsp*. Vil3.

## RESULTS AND DISCUSSION

### Sequence analysis of seven *Thermus* ligase genes

Amino acid sequence alignment of five Gram-negative bacterial NAD<sup>+</sup>-dependent DNA ligases indicates that *Tth* ligase is 93% identical to *T.scotoductus* ligase, 49% to *Rhodothermus marinus* ligase, 48% to *E.coli* ligase and 38% to *Zymomonas mobilis* based on sequence data retrieved from GenBank. Degenerate primers corresponding to highly conserved regions of these ligases were used to amplify fragments of ligase genes from seven *Thermus* strains which represent a worldwide collection: *T.flavus* from Japan, *T.aquaticus* YT-1 and *Thermus* sp. AK16D from Yellowstone National Park in the USA, *T.filiformis* Tok4A2 and *T.filiformis* Tok6A1 from New Zealand, *Thermus* sp. SM32 from the Azores and *Thermus* sp. Vil3 from Portugal. The sequences of amplified ligase fragments ranging from 1.4 to 1.6 kb were determined by directly sequencing the PCR products using an ABI 373 automated sequencer. *Thermus* ligases, in general, were highly conserved during evolution as demonstrated by 85–98% sequence identity. In contrast, the amino acid sequences of the restriction endonuclease *TaqI* and its isoschizomers from the identical strains show only 50–70% amino acid identities (38). *Thermus* ligases in general show 30–40% sequence identities as compared with DNA ligases from other bacteria. The sequence divergence is slightly higher among the different geographic groups than within the same group, which may reflect random drift or adaptation to their respective local environments (Fig. 1). *Thermus flavus*, *T.filiformis* Tok4A2, *T.filiformis* Tok6A1, *Thermus* sp. SM32, *Thermus* sp. Vil3, *T.aquaticus* YT-1 and *Thermus* sp.



**Scheme 1.**



AK16D ligases shared 98.2, 89.9, 89.5, 89.8, 88.3, 88.2 and 88.1% with *T.thermophilus* HB8 DNA ligase, respectively. The adenylation site of the enzymes ( $^{118}\text{KXDG}$ , where X is in general a hydrophobic residue), as identified by site-directed mutagenesis of *Tth* DNA ligase, is completely identical among all *Thermus* ligases; furthermore, the flanking sequences of the adenylation motif are also identical except for *Tsp.* AK16D in which the amino acid residue  $^{117}\text{H}$  before  $^{118}\text{K}$  is substituted by  $^{118}\text{R}$  (Fig. 1B). In non-*Thermus*  $\text{NAD}^+$ -dependent ligases discovered to date, the corresponding position is either a Pro or a Leu. The two isolates from Japan can be distinguished from the other *Thermus* strains by a three amino acid insertion at position 234.

### Cloning, expression and purification of DNA ligase from *Tsp.* AK16D

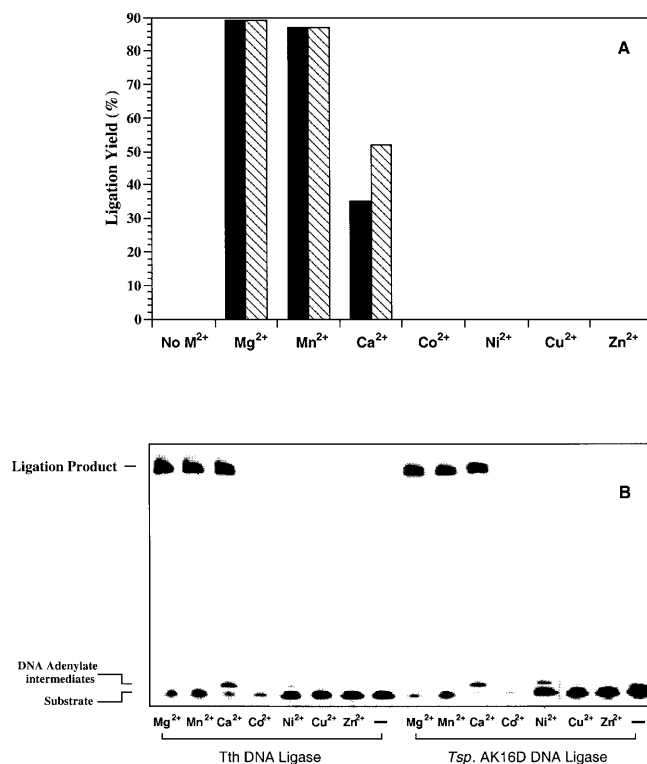
To maximize our chance of finding a *Thermus* ligase with novel properties, we chose *Tsp.* AK16D ligase which showed the least sequence identity as compared with *T.thermophilus* ligase. To obtain the complete sequence of the ORF, the fragments of the N- and C-terminus of the gene were amplified by inverse PCR and were subjected to direct sequencing. The complete ORF of the *Thermus* sp. AK16D ligase gene consists of 674 amino acids, as compared with 676 amino acids for *Tth* ligase (GenBank accession no. 118780) and 674 amino acids for *T.scotoductus* ligase (GenBank accession no. 1085749). The full-length *Thermus* sp. AK16D ligase gene was PCR amplified using *Pfu* polymerase and cloned into expression plasmid pET11c (Novagen). The integrity of the insert containing the ligase gene was verified by DNA sequencing. The pET11c plasmid expressing *Tsp.* AK16D ligase was transformed into competent *E.coli* cells Nova-Blue(DE3)pLysS. Production of ligases was induced by adding IPTG to 1 mM final concentration. *Tsp.* AK16D ligase protein was expressed to ~10% of total cellular proteins. Heating at  $70^\circ\text{C}$  for 15 min denatured most of the *E.coli* proteins while leaving the thermostable ligases as the dominant band. Cibacron blue-based affinity chromatography (Pharmacia) further removed residual *E.coli* proteins and nucleic acids, yielding apparently homogenous *Tsp.* AK16D ligase protein as judged by Coomassie staining.

### Salt, pH and $\text{NAD}^+$ dependence of the ligation reaction

The pH dependence curves of *Tth* ligase and *Tsp.* AK16D ligase are essentially superimposable (data not shown). The optimal pH is 8.5 for both *Tth* ligase and *Tsp.* AK16D ligase with >80% activity observed between pH 7.8 and 9.5. The identity of the pH effect suggests that both of the ligases possess similar local environments at their catalytic centers, which is in agreement with the degree of sequence conservation between the two ligases. The optimum KCl concentrations for *Tth* ligase and *Tsp.* AK16D ligase are 100 and 50 mM, respectively. The optimum  $\text{NAD}^+$  concentration is 1 mM for both *Tth* ligase and *Tsp.* AK16D ligase. The similarity of the  $\text{NAD}$  profiles is in keeping with the highly conserved nature of the N-terminal domain of the ligases which is involved in  $\text{NAD}^+$  binding (22).

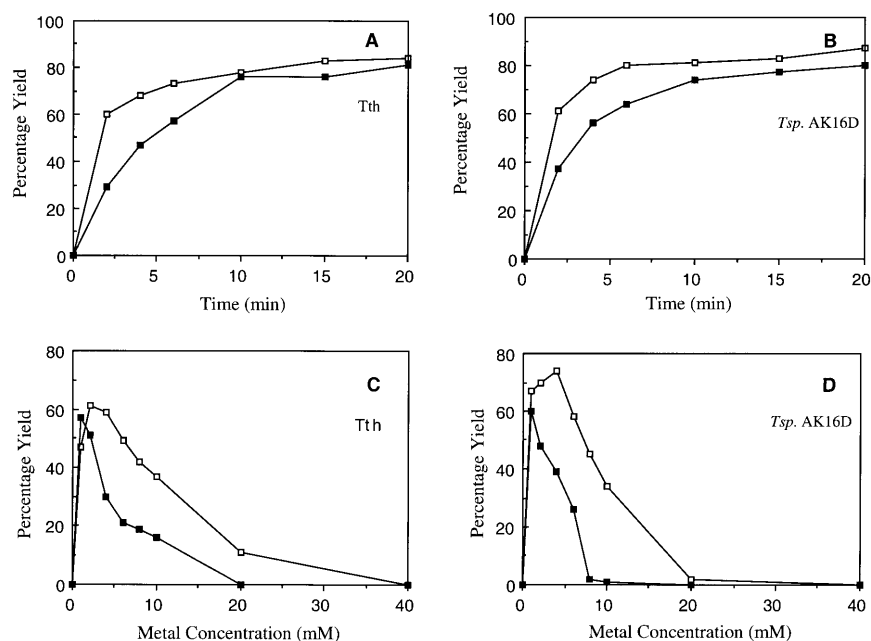
### Effects of divalent metals on the ligation reaction

Divalent metal ion is indispensable for each of the three steps in a ligation reaction: (i) adenylation of a lysine residue in the adenylation motif KXDG; (ii) transfer of the adenylate to the 5'-phosphate to form a DNA-adenylate intermediate; (iii) formation



**Figure 2.** Divalent cation dependence of *Thermus* ligase activity. (A) Ligation reactions with different divalent ions as the metal cofactor. Striped bars, *Tsp.* AK16D ligase; solid bars, *Tth* ligase. Reaction mixtures (20  $\mu\text{l}$ ) containing 20 nM nicked duplex substrate, 0.5 nM *Tth* ligase or 1 nM *Tsp.* AK16D ligase, 20 mM Tris-HCl (pH 8.5), 100 mM KCl (50 mM for *Tsp.* AK16D ligase), 10 mM DTT, 1 mM  $\text{NAD}^+$  and 20  $\mu\text{g/ml}$  BSA and 5 mM indicated divalent cation were incubated at  $65^\circ\text{C}$  for 60 min. (B) Chromatogram of a representative GeneScan gel illustrating ligation product and DNA adenylate intermediate. -, negative control reactions in which ligase was omitted.  $\text{Co}^{2+}$  may have caused precipitation of DNA substrate which resulted in disappearance of the unreacted substrate. Reaction mixtures (20  $\mu\text{l}$ ) containing 20 nM nicked duplex substrate, 1 nM *Tth* ligase or 2 nM *Tsp.* AK16D ligase and 5 mM indicated divalent cation were incubated at  $65^\circ\text{C}$  for 20 min with  $\text{Mg}^{2+}$  or  $\text{Mn}^{2+}$  as metal cofactor. Incubation time with other metal ions was extended to 90 min due to slow ligation rates.

of a phosphodiester bond with the release of adenosine monophosphate (AMP). In general,  $\text{Mg}^{2+}$  is the preferred metal ion for both ATP-dependent and  $\text{NAD}^+$ -dependent ligases. We substituted  $\text{Mg}^{2+}$  with alkaline earth metal ion  $\text{Ca}^{2+}$  and commonly studied period 4 transition metal ions. *Tth* and *Tsp.* AK16D ligases could use  $\text{Mn}^{2+}$  as an alternative metal cofactor to support ligation activity (Fig. 2). Both enzymes were less active with  $\text{Ca}^{2+}$ , while  $\text{Co}^{2+}$ ,  $\text{Ni}^{2+}$ ,  $\text{Cu}^{2+}$  and  $\text{Zn}^{2+}$  failed to support ligation. In comparison, ATP-dependent ligase from *Haemophilus influenzae* uses only  $\text{Mg}^{2+}$  and  $\text{Mn}^{2+}$  as the metal cofactor for nick closure but not  $\text{Ca}^{2+}$ ,  $\text{Co}^{2+}$ ,  $\text{Cu}^{2+}$  and  $\text{Zn}^{2+}$  (14); ATP-dependent ligase from *Chlorella* virus PBCV-1 can use  $\text{Mg}^{2+}$ ,  $\text{Mn}^{2+}$  and  $\text{Co}^{2+}$  but not  $\text{Ca}^{2+}$ ,  $\text{Cu}^{2+}$  and  $\text{Zn}^{2+}$  (4). Using  $\text{Ca}^{2+}$  as the metal cofactor, *Thermus* enzymes were able to convert most of the substrate into the DNA-adenylate intermediate. However, the nick closure activity was reduced which led to accumulation of the DNA-adenylate intermediate (Fig. 2B). A small amount of the intermediate was observed with  $\text{Ni}^{2+}$ , however, ligation product was not observed at the current detection level suggesting that  $\text{Ni}^{2+}$  could not



**Figure 3.** Time course of *Tth* (A) and *Tsp. AK16D* (B) ligase activity in the presence of  $Mg^{2+}$  (open square) or  $Mn^{2+}$  (closed square). Reactions were performed in 100  $\mu$ l mixture as described in Figure 2 legend. Aliquots (5  $\mu$ l) were removed at the indicated time and reactions stopped by adding an equal volume of stop solution. Divalent cation concentration dependence of *Tth* (C) and *Tsp. AK16D* (D) ligase activity.  $Mg^{2+}$ , open square;  $Mn^{2+}$ , closed square. Reactions were performed in 20  $\mu$ l mixture containing 20 nM nicked duplex substrate, 0.5 nM *Tth* ligase or 1 nM *Tsp. AK16D* ligase and indicated concentrations of  $Mg^{2+}$  or  $Mn^{2+}$  at 65 °C for 2 min.

support the nick closure step (Fig. 2B). Similar DNA–adenylate intermediates were observed with T4 and ATP ligases when non-matched substrates were used (5,23). To further compare the relative activities of the two *Thermus* ligases with  $Mg^{2+}$  and  $Mn^{2+}$ , we first monitored the generation of ligation product over a 20 min time period. The *Thermus* enzymes were consistently more active with  $Mg^{2+}$  than with  $Mn^{2+}$  (Fig. 3A and B). Second, we assayed ligation activity with up to 40 mM  $Mg^{2+}$  or  $Mn^{2+}$  (Fig. 3C and D). Both of the enzymes responded sensitively to the change of the metal ion concentration in the reaction mixture. At high  $M^{2+}$  concentrations, the high ionic strength may inhibit the enzyme activity, consistent with the KCl dependence profile. Similar to the time course results, the *Thermus* enzymes were more active with  $Mg^{2+}$  than with  $Mn^{2+}$ . The discrepancy in the relative activity of *Thermus* ligases between our study and an earlier report may be due to our use of cloned enzymes while the earlier work used purified native enzyme (41).

### Steady-state kinetics

The steady-state kinetic constants were measured by monitoring the formation of fluorescently labeled ligation product over time using substrate concentrations spanning estimated  $K_m$  values (Table 1). The steady-state properties of *Tsp. AK16D* ligase were similar to *Tth* ligase, indicating that the catalytic channels are highly conserved in *Thermus* ligases. The average  $K_m$  value of ~90 nM for *Thermus* ligases is similar to the  $K_m$  value of 50 nM for *E.coli* ligase (42) and ~10-fold higher than vaccinia virus ATP-dependent ligase (21). The average  $k_{cat}$  value of ~45 turnovers/min for *Thermus* ligases is higher than the  $k_{cat}$  value of 28 turnovers/min for *E.coli* ligase (42).

**Table 1.** Steady-state kinetics of *Tth* and *Tsp. AK16D* ligases<sup>a</sup>

Ligase	$K_m$ (nM)	$k_{cat}$ (per min)	$k_{cat}/K_m$ (per M/s)
<i>Tth</i>	87	56	$1.1 \times 10^7$
<i>Tsp.AK16D</i>	104	38	$0.62 \times 10^7$

<sup>a</sup>Reaction conditions were described in the Figure 2 legend. Results represent the average of at least three experiments.

### Ligation of gapped or inserted DNA duplex substrates

Gapped substrates were formed by deleting 1 or 2 nt from the 3'-hydroxyl site of oligonucleotide LP3'C and inserted substrates were formed by adding 1 or 2 nt at the 3'-hydroxyl site of oligonucleotide LP3'C (sequences in Materials and Methods). Gapped or inserted duplexed DNA sequences are distinctively different from normal nicked substrate. Under our experimental conditions, no ligation was detectable with 1 or 2 nt gapped or 2 nt insertion substrates for either *Tth* or *Tsp. AK16D* ligase. As for 1 nt insertion substrates, only an A insertion gave a trace amount of ligated products for both ligases. All other 1 nt insertions at the ligation junction could not be ligated. In contrast, *H.influenzae* ligase and *Chlorella* ligase demonstrate observable ligation with a 1 nt gap (4,14). In the case of vaccinia ligase, the ligation of a 1 nt gap is negligible but the formation of DNA–adenylate intermediate is significant, suggesting that the major impact of using 1 nt gapped substrate is on nick closure (5). We did not observe formation of DNA–adenylate intermediate with the *Thermus* enzymes, suggesting that most of the gapped or inserted substrates may have abolished the possibility of completing the second step in the ligation cycle, adenylation of DNA substrate at the 5'-phosphate.

### Thermus DNA ligase fidelity

*Tth* DNA ligase is more discriminative when the mismatch is located at the 3'-side of the nick junction. 3'-G/T or 3'-T/G is the only mismatch that shows observable mismatch ligation (26). To evaluate the fidelity of the cloned *Tsp.* AK16D ligase, we compared the rate ratio of match over 3'-T/G mismatch ligation with wild-type and K294R mutant *Tth* DNA ligases along with T4 ligase from a commercial source (Table 2). T4 ligase demonstrated high catalytic efficiency toward both matched and 3'-T/G mismatched substrate such that a ligation fidelity of 50 was obtained. *Thermus* ligases appeared to be less efficient in match ligation as evidenced by the requirement for higher enzyme concentration to achieve comparable match ligation rate. However, under the same assay conditions, *Thermus* enzymes were far less prone to ligate a 3'-T/G mismatch. As a result, the fidelity of *Thermus* enzymes was 17- to 126-fold higher than T4 ligase (Table 2, Ligation fidelity 1). The fidelity of the newly cloned *Tsp.* AK16D ligase was similar to K294R *Tth* mutant but 6-fold higher than the wild-type *Tth* enzyme. A DNA-adenylate intermediate was observed with 3'-T/G mismatch ligation (data not shown), suggesting that a mismatch at the 3' ligation junction imposes substantial constraints on the ability of *Thermus* ligases to close the nick, thereby limiting the turnover of DNA-adenylate intermediate into ligated product and free AMP (the third step of the ligation cycle). We further examined the effects of moving the T/G mismatch 1 bp away from the ligation junction. The rates of ligation with a T/G mismatch at the penultimate 3'-end in general improved several-fold as compared with the T/G mismatch at the 3'-end of the ligation junction. However, the ligation rates were still much slower than those of match ligation, emphasizing the importance of nucleotide complementarity near the ligation junction as well as the ultimate critical role of the perfect base pair at the 3'-end in controlling the ligation reaction. Consequently, the ligation fidelity when the mismatch was at the second position from the 3'-side (Table 2, Ligation fidelity 2) was lower than that when the mismatch was located immediately at the ligation

junction. It is noteworthy that the *Tsp.* AK16D enzyme maintains extremely high fidelity ( $1.1 \times 10^3$ ) even when the mismatch is at the penultimate position, further underscoring the discriminative power of this new *Thermus* ligase.

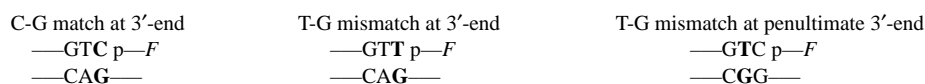
### Thermostable DNA ligase fidelity in the presence of Mn<sup>2+</sup>

Many enzymes such as DNA polymerase and restriction endonucleases demonstrate relaxed specificity when Mn<sup>2+</sup> was used as the metal cofactor. The influence of metal ion substitution on ligase fidelity has not been fully investigated although it is known Mn<sup>2+</sup> can be used as an alternative metal cofactor for the ligation reaction (4,14; this work). We determined the reaction rates of the match and mismatch ligation for *Tsp.* AK16D ligase and *Tth* ligase. As shown in Table 3, the match ligation rates were higher with Mg<sup>2+</sup> than with Mn<sup>2+</sup> (Table 2), in agreement with the consistent high ligation rate under various Mg<sup>2+</sup> conditions (Fig. 3). The mismatch ligation rate of *Tth* ligase was ~6-fold higher with Mn<sup>2+</sup> than with Mg<sup>2+</sup> while that of *Tsp.* AK16D ligase was ~4-fold higher. Thus, as with other previously studied DNA enzymes, DNA ligases also demonstrate relaxed specificity when Mg<sup>2+</sup> is substituted with Mn<sup>2+</sup>. As a result, the fidelity factors of *Tth* ligase and *Tsp.* AK16D ligase were reduced 12- and 6-fold, respectively (Table 2). Remarkably, the *Tsp.* AK16D enzyme retains 12-fold higher fidelity against mismatch ligation than the *Tth* enzyme. In contrast to using Mg<sup>2+</sup> as the metal cofactor, *Tth* ligase did not generate DNA-adenylate intermediate during 3'-T/G mismatch ligation with Mn<sup>2+</sup> (data not shown). This observation suggests that the nick closure of a 3'-T/G mismatch by the *Tth* enzyme is accelerated with Mn<sup>2+</sup>. On the other hand, the *Tsp.* AK16D enzyme accumulated DNA-adenylate intermediate during 3'-T/G mismatch ligation with either Mg<sup>2+</sup> or Mn<sup>2+</sup> (data not shown). These results indicate that the nick closure of a 3'-T/G mismatch with Mn<sup>2+</sup> by *Tsp.* AK16D DNA ligase remains as the rate-limiting step, which accounts for the higher fidelity of this enzyme.

Table 2. DNA ligase fidelity<sup>a</sup>

Ligase	Enzyme concentration (nM)	Initial rates of C-G match (fmol/min)	Initial rates of T-G mismatch at 3'-end (fmol/min)	Initial rates of T-G mismatch at penultimate 3'-end (fmol/min)	Ligation fidelity 1 <sup>b</sup>	Ligation fidelity 2 <sup>c</sup>
T4	0.5	$1.4 \times 10^2$	2.8	7.1	$5.0 \times 10$	$1.9 \times 10$
<i>Tth</i> -wt	1.25	$5.5 \times 10$	$6.5 \times 10^{-2}$	$2.9 \times 10^{-1}$	$8.4 \times 10^2$	$1.9 \times 10^2$
<i>Tth</i> -K294R	12.5	$1.5 \times 10^2$	$2.3 \times 10^{-2}$	$4.3 \times 10^{-1}$	$6.3 \times 10^3$	$3.4 \times 10^2$
<i>Tsp.</i> AK16D	12.5	$1.3 \times 10^2$	$2.5 \times 10^{-2}$	$1.2 \times 10^{-1}$	$5.1 \times 10^3$	$1.1 \times 10^3$

<sup>a</sup>T4 DNA ligase fidelity was assayed at 37°C, thermophilic ligase fidelity was assayed at 65°C (details in Materials and Methods). A schematic illustration of matched and mismatched substrates are as follows:



<sup>b</sup>Ligation fidelity 1, initial rate of C-G match/initial rate of T-G mismatch at the 3'-end.

<sup>c</sup>Ligation fidelity 2, initial rate of C-G match/initial rate of T-G mismatch at the penultimate 3'-end.

The concentrations of DNA ligases used in each experiment are as indicated. Results were calculated as the average of at least two experiments.

**Table 3.** DNA ligase fidelity with Mn<sup>2+</sup><sup>a</sup>

Ligase	Concentration (nM)	Initial rate of C-G match (fmol/min)	Initial rate of T-G mismatch (fmol/min)	Ligation fidelity
<i>Tth</i> -wt	1.25	2.6 × 10	3.7 × 10 <sup>-1</sup>	7.0 × 10
<i>Tsp</i> . AK16D	12.5	9.5 × 10	1.1 × 10 <sup>-1</sup>	8.6 × 10 <sup>2</sup>

<sup>a</sup>Reaction conditions were identical to in Table 2, except that 10 mM Mn<sup>2+</sup> was used in place of Mg<sup>2+</sup>. Ligation fidelity was defined as the ratio of initial rate of C-G match divided by initial rate of T-G mismatch at the 3'-end. Results were calculated as the average of at least two experiments.

### Concluding remarks

Studies on *Tth* DNA ligase has deepened our understanding of thermostable ligases and has reaffirmed the common theme of ligation, adenylation of ligase at the KXDG motif (18). This study reveals that *Thermus* ligases may differ from each other as to substrate specificity despite their highly identical primary protein sequences. A highly homologous structure can be anticipated from various *Thermus* ligases, but subtle local environments may dictate the probability of accepting a particular mismatch as the substrate. The fidelity of the *Thermus* ligases may be determined by multiple domains, multiple motifs and/or multiple sequence elements. In a comparison of *Tth* and *Tsp*. AK16D ligases, one can find that although K294R (in an identical local environment; Fig. 1B) enhances the fidelity of *Tth* ligase (26; this work), *Tsp*. AK16D ligase with a K in this position can still demonstrate superior mismatch discrimination. Additional sequence elements remain to be uncovered. We do not yet know whether the R substitution at the adjacent position to the KXDG motif has any effects on the *Tsp*. AK16D ligase specificity, but studies on *Chlorella* ligase has emphasized the importance of occupying the AMP-binding pocket for nick recognition (43). The accumulation of DNA-adenylate intermediate with some divalent metal ions by *Tsp*. AK16D ligase asserts that the nick closure step of a ligation reaction can be sensitive to the selection of metal ions, gapped substrates and mismatched substrates. More structural and functional studies on *Tsp*. AK16D ligase could reveal how this enzyme achieves high fidelity with different substrates and different metal ions.

### ACKNOWLEDGEMENTS

We thank Jing Lu for *Thermus* genomic DNA preparation and Jianying Luo and Marilyn Khanna for protein purification. We also thank Dr Dale B. Wigley for communicating results prior to publication. This work was supported by grants from the National Institutes of Health (GM-41337-09 and PO1-CA65930-02-04) and the Applied Biosystems Division of Perkin Elmer Corp.

### REFERENCES

- Lehman, I.R. (1974) *Science*, **186**, 790–797.
- Dunn, J.J. and Studier, F.W. (1981) *J. Mol. Biol.*, **148**, 303–330.
- Weiss, B. and Richardson, C.C. (1967) *Proc. Natl Acad. Sci. USA*, **57**, 1021–1028.
- Ho, C.K., Van Etten, J.L. and Shuman, S. (1997) *J. Virol.*, **71**, 1931–1937.
- Shuman, S. (1995) *Biochemistry*, **34**, 16138–16147.
- Kletzin, A. (1992) *Nucleic Acids Res.*, **20**, 5389–5396.
- Bult, C.J., White, O., Olsen, G.J., Zhou, L., Fleischmann, R.D., Sutton, G.G., Blake, J.A., Fitz Gerald, L.M., Clayton, R.A., Gocayne, J.D., Kerlavage, A.R., Dougherty, B.A., Tomb, J.F., Adams, M.D., Reich, C.I., Overbeek, R., Kirkness, E.F., Weinstock, K.G., Merrick, J.M., Glodek, A., Scott, J.L., Geoghagen, N., Weidman, J.F., Fuhrmann, J.L., Venter, J.C. *et al.* (1996) *Science*, **273**, 1058–1073.
- Andaluz, E., Larriba, G. and Calderone, R. (1996) *Yeast*, **12**, 893–898.
- Ramos, W., Tappe, N., Talamantez, J., Friedberg, E.C. and Tomkinson, A.E. (1997) *Nucleic Acids Res.*, **25**, 1485–1492.
- Schar, P., Herrmann, G., Daly, G. and Lindahl, T. (1997) *Genes Dev.*, **11**, 1912–1924.
- Tomkinson, A.E. and Levin, D.S. (1997) *Bioessays*, **19**, 893–901.
- Tomkinson, A.E. and Mackey, Z.B. (1998) *Mutat. Res.*, **407**, 1–9.
- Wang, Y.C., Burkhart, W.A., Mackey, Z.B., Moyer, M.B., Ramos, W., Husain, I., Chen, J., Besterman, J.M. and Tomkinson, A.E. (1994) *J. Biol. Chem.*, **269**, 31923–31928.
- Cheng, C. and Shuman, S. (1997) *Nucleic Acids Res.*, **25**, 1369–1374.
- Deckert, G., Warren, P.V., Gaasterland, T., Young, W.G., Lenox, A.L., Graham, D.E., Overbeek, R., Snead, M.A., Keller, M., Aujay, M., Huber, R., Feldman, R.A., Short, J.M., Olsen, G.J. and Swanson, R.V. (1998) *Nature*, **392**, 353–358.
- Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirkness, E.F., Kerlavage, A.R., Bult, C.J., Tomb, J.F., Dougherty, B.A., Merrick, J.M. *et al.* (1995) *Science*, **269**, 496–512.
- Shuman, S. and Ru, X.M. (1995) *Virology*, **211**, 73–83.
- Luo, J. and Barany, F. (1996) *Nucleic Acids Res.*, **24**, 3079–3085.
- Doherty, A.J., Ashford, S.R. and Wigley, D.B. (1996) *Nucleic Acids Res.*, **24**, 2281–2287.
- Subramanya, H.S., Doherty, A.J., Ashford, S.R. and Wigley, D.B. (1996) *Cell*, **85**, 607–615.
- Sekiguchi, J. and Shuman, S. (1997) *Nucleic Acids Res.*, **25**, 727–734.
- Singleton, M.R., Håkansson, K., Timson, D.J. and Wigley, D.B. (1999) *Structure*, in press.
- Wu, D.Y. and Wallace, R.B. (1989) *Gene*, **76**, 245–254.
- Husain, I., Tomkinson, A.E., Burkhart, W.A., Moyer, M.B., Ramos, W., Mackey, Z.B., Besterman, J.M. and Chen, J. (1995) *J. Biol. Chem.*, **270**, 9683–9690.
- Rossi, R., Montecucco, A., Ciarrocchi, G. and Biamonti, G. (1997) *Nucleic Acids Res.*, **25**, 2106–2113.
- Luo, J., Bergstrom, D.E. and Barany, F. (1996) *Nucleic Acids Res.*, **24**, 3071–3078.
- Wu, D.Y. and Wallace, R.B. (1989) *Genomics*, **4**, 560–569.
- Landegren, U., Kaiser, R., Sanders, J. and Hood, L. (1988) *Science*, **241**, 1077–1080.
- Landegren, U. (1993) *Bioessays*, **15**, 761–765.
- Barany, F. (1991) *PCR Methods Appl.*, **1**, 5–16.
- Barany, F. (1991) *Proc. Natl Acad. Sci. USA*, **88**, 189–193.
- Day, D.J., Speiser, P.W., White, P.C. and Barany, F. (1995) *Genomics*, **29**, 152–162.
- Khanna, M., Park, P., Zirvi, M., Cao, W., Picon, A., Day, J., Paty, P. and Barany, F. (1999) *Oncogene*, in press.
- Thorbjarnardottir, S.H., Jonsson, Z.O., Andresson, O.S., Kristjansson, J.K., Eggertsson, G. and Palsdottir, A. (1995) *Gene*, **161**, 1–6.
- Jonsson, Z.O., Thorbjarnardottir, S.H., Eggertsson, G. and Palsdottir, A. (1994) *Gene*, **151**, 177–180.
- Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Applied Biosystems Inc. (1992) *The Complete Guide—Evaluating and Isolating Synthetic Oligonucleotides*. Applied Biosystems Inc., Foster City, CA.
- Cao, W., Lu, J. and Barany, F. (1997) *Gene*, **197**, 205–214.
- Wetmur, J.G., Wong, D.M., Ortiz, B., Tong, J., Reichert, F. and Gelfand, D.H. (1994) *J. Biol. Chem.*, **269**, 25928–25935.
- Higgins, D.G. and Sharp, P.M. (1989) *Comput. Appl. Biosci.*, **5**, 151–153.
- Takahashi, M., Yamaguchi, E. and Uchida, T. (1984) *J. Biol. Chem.*, **259**, 10041–10047.
- Modrich, P., Anraku, Y. and Lehman, I.R. (1973) *J. Biol. Chem.*, **248**, 7495–7501.
- Sriskanda, V. and Shuman, S. (1998) *Nucleic Acids Res.*, **26**, 525–531.