JOURNAL OF FERMENTATION AND BIOENGINEERING
Vol. 76, No. 4, 265–269. 1993

# Cloning, Nucleotide Sequence, and Expression in *Escherichia coli* of DNA Polymerase Gene (*polA*) from *Thermus thermophilus* HB8

KEIKO ASAKURA,[1] HIDEYUKI KOMATSUBARA,[2] SHIRO SOGA,[1] TETSUYA YOMO,[1] MASANORI OKA,[2] SHIGENORI EMI,[2] AND ITARU URABE[1]*

*Department of Biotechnology, Faculty of Engineering, Osaka University, 2–1 Yamadaoka, Suita, Osaka 565[1] and Toyobo Tsuruga Institute of Biotechnology, 10–24 Toyo-cho, Tsuruga, Fukui 914,[2] Japan*

Received 13 May 1993/Accepted 29 June 1993

**A gene coding for a thermostable DNA polymerase I (Tth Pol I) was cloned from *Thermus thermophilus* HB8, and its nucleotide sequence was identified. The Tth Pol I gene (*polA*) has an open reading frame of 2505 base pairs available to encode a peptide of 834 amino acids. Another incomplete open reading frame was also found upstream from the *polA* gene. In the deduced amino acid sequence of Tth Pol I, which shows 87% similarity to that of *Thermus aquaticus* DNA polymerase, there is the leucine zipper structure from Leu458 to Leu486. The Tth Pol I gene was subcloned in a high-expression vector. *Escherichia coli* cells harboring the hybrid plasmid produced about 100,000 units of thermostable DNA polymerase in a 200-ml culture.**

DNA polymerase from thermophilic bacteria is highly useful in the polymerase chain reaction (PCR) method (1). Previously, a DNA polymerase (Taq Pol I) was purified from *Thermus aquaticus* (2), and its gene was cloned, sequenced, and expressed in *Escherichia coli* (3). Another thermostable DNA polymerase I (Tth Pol I) was purified from *Thermus thermophilus* HB8, and has been shown to be useful in the PCR method (unpublished results). In this work, we cloned the Tth Pol I gene (*polA*) and expressed it in *E. coli* cells. The nucleotide sequence of the *polA* gene was identified and compared with those of other known genes. The cloned gene and the production system will be useful for the study and application of this enzyme.

## MATERIALS AND METHODS

**Bacterial strains and plasmids** *T. thermophilus* HB8 (4), a generous gift from Dr. Tairo Oshima (Tokyo Institute of Technology), was used as the source of Tth Pol I and its gene. The following bacterial strains and plasmids were used: *E. coli* strains JM109 (*recA1* Δ(*lac-pro*) *endA1 gyrA96 thi-1 hsdR17 supE44 relA1*, F′ *traD36 proAB*) (5), and KP3998 ($F^-$ *hsdS20* ($r_B^-$ $m_B^-$) *ara-14 proA2 lacI*$^q$ *galK2 rpsL20 xyl-5 mlt-1 supE44* $\lambda^-$) (6); and plasmids pUC 18 (7), pUC19 (5, 7), and pKP1500 (6). *E. coli* KP3998 and pKP1500 were a generous gift from Dr. Takeyoshi Miki (Kyushu University). *T. thermophilus* was grown at 75°C on a medium containing tryptone (20 g), yeast extract (10 g), and NaCl (10 g) in 1 *l* (pH 7.2); *E. coli* was grown at 37°C on Luria-Bertani medium (8) containing ampicillin (50 μg/ml, when necessary).

**Enzymes and chemicals** Purified Tth Pol I, restriction endonucleases, and other enzymes for DNA manipulation were obtained from Toyobo Co. Ltd. (Osaka). Taq Pol I was purchased from Takara Shuzo Co. Ltd. (Kyoto).

**DNA manipulation and transformation** Preparation of plasmid DNA, enzyme reactions, and transformation of *E. coli* cells was carried out as described by Maniatis *et al.* (8).

**Preparation of primers and probe DNA** Partial amino acid sequences of Tth Pol I were identified by automated Edman sequencing of peptide fragments obtained by digestion with *Achromobacter* protease I. To prepare probe DNAs by the PCR method (1), the following DNA primers were prepared: primer 1 (5′-ATGAGGGGGATGCTGCCCCTCTTTGAG-3′), encoding the N-terminal sequence of Taq Pol I (3); primer 2 (5′-GACATCCACACGCAGACCGCCAGCTGGATGTTC-3′), encoding an amino acid sequence of Tth Pol I identified in this study (Asp-Ile-His-Thr-Gln-Thr-Ala-Ser-X-Met-Phe), which is almost the same as that of Taq Pol I (Asp637-Ile-His-Thr-Glu-Thr-Ala-Ser-Trp-Met-Phe) (3); primer 3, the anti-sequence of primer 2; primer 4 (5′-GTCGGCGGCGGTGCCCTG-3′), the anti-sequence of the Taq Pol I sequence (Gln754-Gly-Thr-Ala-Ala-Asp) which is identical to the *E. coli* Pol I sequence (3); and primer 5 (5′-TCACTCCTTGGCGGAGAGCCA-3′), the anti-sequence of the Taq Pol I sequence that encoding its C-terminal region including the stop codon. Among these primers, two pairs (primers 2+4 and primers 2+5) were able to amplify DNA fragments (about 370 and 570 base pairs (bp), respectively) which seemed to originate from the Tth Pol I gene under the following PCR conditions: a reaction mixture (0.1 ml) containing 0.3 μg genomic DNA, 1 μM each primer, 0.2 mM each dNTP (dATP, dCTP, dTTP, dGTP), 10 mM Tris-HCl (pH 8.8), 1.5 mM $MgCl_2$, 0.1 mg/ml gelatin, and 2.5 units Taq Pol I, was incubated for 30 cycles of 2 min at 92°C, 3 min at 50°C, and 2 min at 72°C using the Program Temperature Control System PC-700 (Astec Co. Ltd., Fukuoka).

The 370-bp DNA probe (probe A) was found to be divided into two parts with a similar size by *Hin*dIII digestion. We cloned the two parts separately by the following procedure. The 370-bp DNA fragment was inserted into the *Sma*I site of pUC19, and the resulting hybrid plasmid was cut into two parts by *Hin*dIII digestion. The longer part was self-ligated into the plasmid pUC19-SH, and the shorter one (about 190 bp) was inserted into the

* Corresponding author.

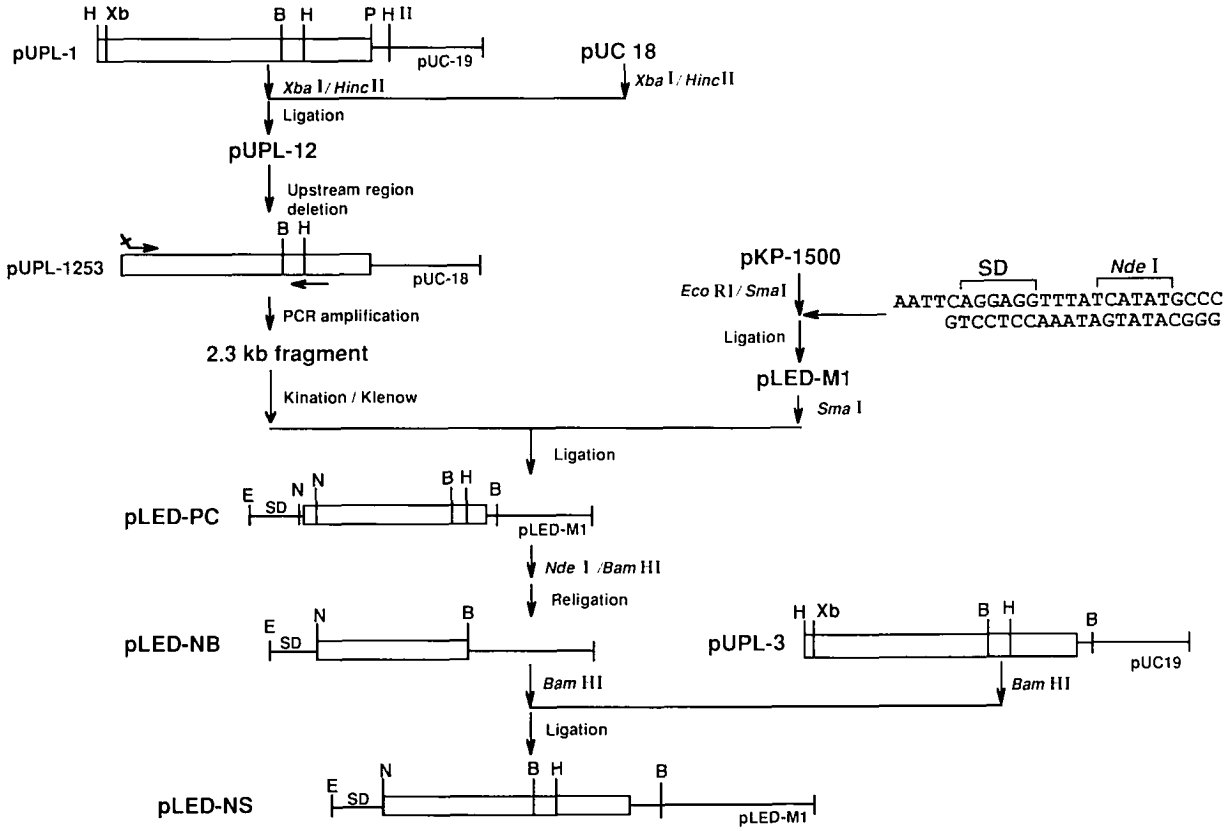


FIG. 1. Scheme for the construction of a Tth Pol I high-expression plasmid, pLED-NS. Restriction sites are abbreviated as follows: B, *Bam*HI; E, *Eco*RI; H, *Hin*dIII; HII, *Hin*cII; N, *Nde*I; P, *Pst*I; Xb, *Xba*I; Xh, *Xho*I.

*Hin*dIII site of pUC19, resulting in the plasmid pUC19-HH. Probe A1 was obtained as the 180-bp *Xba*I-*Hin*dIII fragment of pUC19-SH, and probe A2 was the 180-bp *Hin*dIII-*Pst*I fragment of pUC19-HH. These probes were used for cloning the Tth Pol I gene.

**Cloning of Tth Pol I gene** Chromosomal DNA of *T. thermophilus* HB8 prepared by the method of Saito and Miura (9) was digested with *Hin*dIII. After agarose gel electrophoresis (8) of the digested DNA, fragments of about 2.8 kilo base pairs (kb) and 3.2 kb that hybridized with probe A1 and probe A2, respectively, were recovered from the gel by the method of Vogelstein and Gillespie (10) using a Geneclean kit (Bio101 Inc.). The DNA fragments were ligated with pUC19 that had been digested with *Hin*dIII, and the ligated DNA was introduced into *E. coli* JM109. The plasmid DNA obtained from the ampicillin-resistant transformants were analyzed by hybridization with probe A1 and probe A2, and positive plasmids, p1–100 and p5–7, which hybridized with probe A1 and probe A2, respectively, were obtained. The inserted 2.8-kb *Hin*dIII fragment of p1–100 was ligated with a modified p5–7, which was prepared by digestion of p5–7 with *Pst*I to shorten the inserted fragment from 3.2 kb to 1.5 kb, religation, and then digestion with *Hin*dIII. The hybrid plasmid thus obtained was named pUPL1. The 2.8-kb *Hin*dIII fragment of p1–100 was also ligated with another modified p5–7, which was prepared by digestion with *Xho*I and *Sal*I to shorten the inserted fragment to 0.5 kb, religation, and then digestion with *Hin*dIII. This plasmid was named pUPL3.

**Subcloning of Tth Pol I gene** The following procedures are illustrated in Fig. 1. A 4.3-kb *Xba*I-*Hin*cII fragment of pUPL1 was ligated with pUC18 that had been digested with *Xba*I and *Hin*cII. The resulting hybrid plasmid (pUPL12) was digested with *Kpn*I and *Xba*I, the upstream region of the Tth Pol I gene was shortened by digesting with exonuclease III and then mung bean nuclease. One of the shortened plasmids was named pUPL1253, which was the shortest one containing the whole Tth Pol I gene.

To increase the expression level of the Tth Pol I gene in *E. coli*, the gene was inserted in a high-expression vector, pLED-M1, which was derived from pKP1500 (6) in our laboratory, as shown in Fig. 1. An N-terminal part of the coding region of the Tth Pol I gene was amplified by the PCR method (1) using the primers of 5′-GGCATATGGAGGCGATGCTTCCGCTCTT-3′ and 5′-GTCGGCGGCGGTGCCCTG-3′; by the former primer, the *Nde*I site was introduced overlapping the initiation codon. 2.3-kb fragments of the PCR products were obtained by agarose gel electrophoresis as described above, the 5′-ends were phosphorylated by polynucleotide kinase, and the terminal single-stranded regions of the fragments (if present) were filled in by the Klenow fragment. The 2.3-kb DNA fragment containing the N-terminal part of the Tth Pol I gene was inserted into the *Sma*I site of pLED-M1, and the resulting plasmid (pLED-PC) was digested with *Nde*I, religated, then digested with *Bam*HI and religated again to yield the plasmid pLED-NB. The remaining C-terminal part of the Tth Pol I gene was obtained as the *Bam*HI fragment of pUPL3, and was inserted into the *Bam*HI site of pLED-NB to yield the

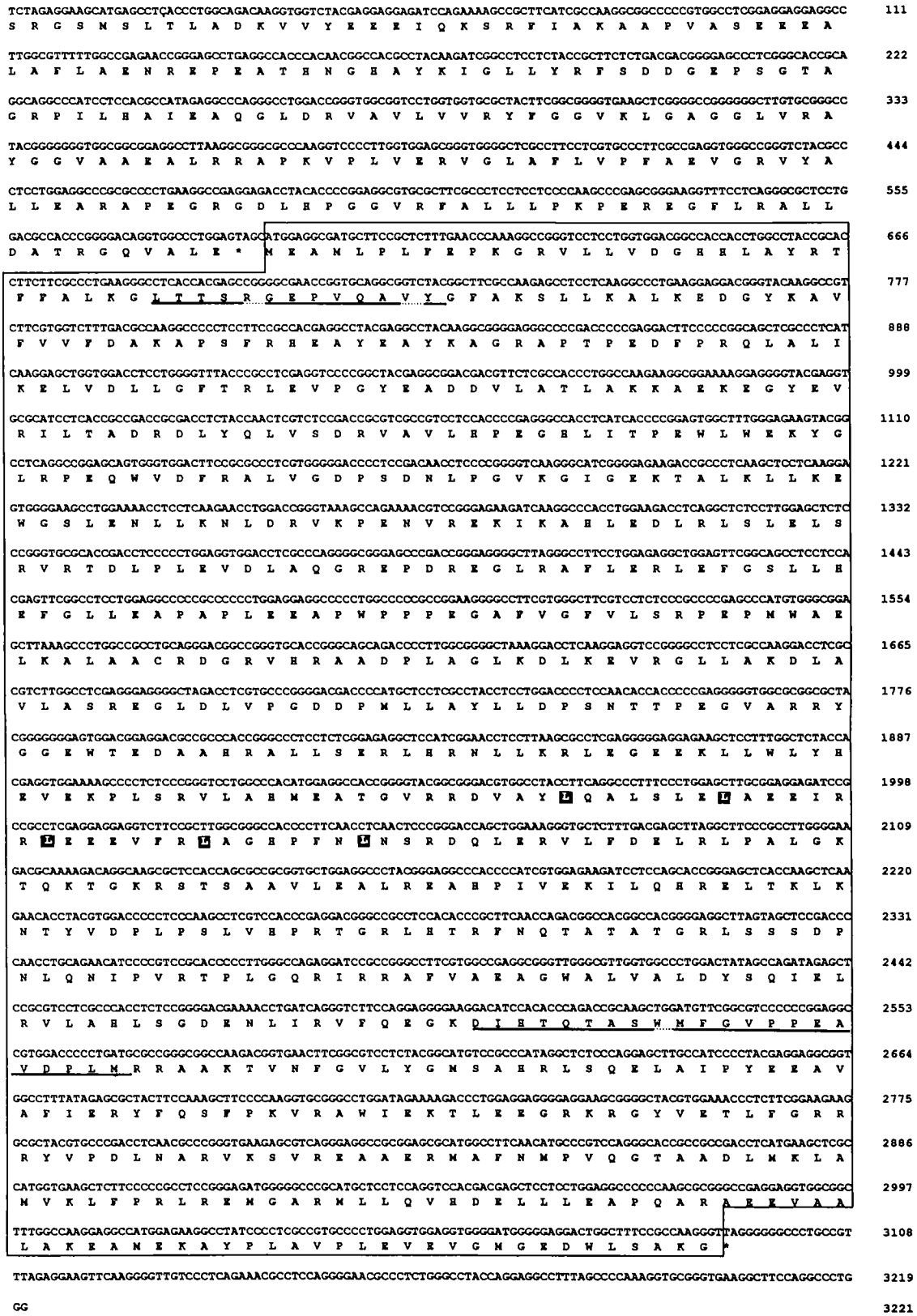

FIG. 2. Nucleotide sequence of the Tth Pol I gene. The nucleotide sequence was identified for both strands. The amino acid sequences deduced from the nucleotide sequence are also shown below. The amino acid sequences, which are identical to the sequences determined by automated Edman sequencing of peptide fragments of the purified enzyme (Ser-*Leu-Thr-Thr-Ser*-X-*Gly-Glu-Pro-Val-Gln-Ala*-X-*Tyr* and *Asp-Ile-His-Thr-Gln-Thr-Ala-Ser*-X-*Met-Phe-Gly-Val-Pro-Pro-Glu-Ala-Val-Asp-Pro-Leu-Met*), are underlined. The region coding for Tth Pol I is boxed, and the five leucine residues forming the leucine zipper structure are outlined in black.

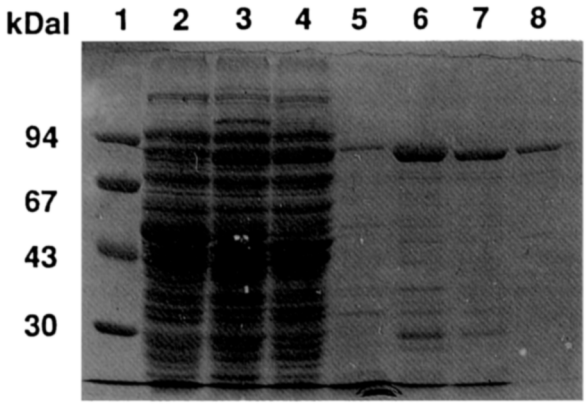


FIG. 3. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis. Electrophoresis was done according to the method of Laemmli (19). The volume of crude cell extract applied was 10 $\mu$l each. Protein bands were stained with Coomassie Brilliant Blue R250. Lane 1, Marker proteins; lanes 2–4, crude cell extract before heat treatment; lanes 5–7, supernatant of heat-treated crude cell extract; lanes 2 and 5, *E. coli* KP3998 harboring pLED-M1 (negative control); lanes 3 and 6, *E. coli* KP3998 harboring pLED-NS; lanes 4 and 7, *E. coli* JM109 harboring pUPL1253; lane 8, purified Tth Pol I (Toyobo; 40 units).

plasmid pLED-NS.

**Preparation of crude enzyme solution** *E. coli* cells obtained from 200-ml culture were suspended in 10 ml 50 mM Tris-HCl (pH 7.5) containing 1 mM EDTA and 2.4 mM phenylmethanesulfonyl fluoride, and disrupted by sonication. The supernatant obtained by centrifugation (crude cell extract) was heated at 75°C for 30 min, centrifuged, and used as a crude enzyme solution.

**Enzyme assay** DNA polymerase activity was measured at 75°C for 30 min in an assay mixture (50 $\mu$l) containing 67 mM Tris-HCl (pH 8.8), 16 mM $(NH_4)_2SO_4$, 6.7 mM $MgCl_2$, 0.2 mM each dATP, dGTP, and dCTP, 0.2 mM [$^3$H]-dTTP (5 $\mu$Ci/mmol), 1 $\mu$g M13mp19 single-stranded DNA, 10 pmol P7 primer, 10 mM mercaptoethanol, 5 mg/ml bovine serum albumin, and enzyme solution (5 $\mu$l). Reactions were stopped by chilling on ice and adding 50 $\mu$l 0.2 M sodium pyrophosphate and 0.1 ml 1 M perchloric acid, and the mixture was then incubated at 0°C for more than 10 min. Precipitated DNA was collected on GF/C filter discs, and washed with 0.1 M HCl (4 ml), then 95% ethanol (2 ml), dried, and counted. One unit of enzyme activity corresponds to 10 nmol of the deoxynucleotides incorporated in 30 min.

## RESULTS AND DISCUSSION

The *polA* gene of *T. thermophilus* was found to be divided into two parts by *Hin*dIII digestion, and we deduced that a 2.8-kb *Hin*dIII fragment and a 3.2-kb *Hin*dIII fragment of the chromosomal DNA were likely to contain the N-terminal and C-terminal halves, respectively, of the gene (data not shown). As there were no other restriction fragments suitable for cloning the *polA* gene, we decided to clone these two *Hin*dIII fragments separately and to assemble the full-length gene. The cloned fragment in pUPL1 was sequenced by the dideoxynucleotide chain-termination method (11), and the results are shown in Fig. 2. There is an open reading frame of 2505 bp available to encode a peptide of 834 amino acids ($M_r$ 94,048). The nucleotide and deduced amino acid sequences of the cloned Tth Pol I gene show high similarity (85% and 87% identities, respectively) to those of the Taq Pol I gene (3). Hence, Tth Pol I is a member of family A DNA polymerases (12), which include *E. coli* DNA polymerase I (Pol I) and Taq Pol I.

In addition to the *polA* gene, there is another incomplete open reading frame which starts beyond the 5′-terminal of the sequence shown in Fig. 1 and ends at the stop codon (TAG) one bp upstream from the initiation codon of the *polA* gene. The deduced amino acid sequence of the incomplete open reading frame, which is also found in the upper region of the Taq Pol I sequence (3), shows 33.8% identity (195 amino acids overlap) to that of a putative X-Pro dipeptidase encoded in the *pepQ* gene of *E. coli* (13); this similarity was confirmed to be significant by the method of Pearson and Lipman (14).

In the amino acid sequence of Tth Pol I (Fig. 2), the leucine zipper structure, a periodic repetition of leucine residues at every seventh position (15), is found from Leu458 to Leu486 showing the periodic array of five leucine residues. The same structure is also found at the corresponding positions in the sequences of Taq Pol I (Leu456 to Leu484) (3) and *E. coli* Pol I (Leu552 to Leu580) (16), though the second Leu is replaced by Val in the structure of Taq Pol I and the fourth Leu is replaced by Ile in that of *E. coli* Pol I. The tertiary structure of the large fragment (Klenow fragment) of *E. coli* Pol I (17) shows that the first three Leu residues are on helix H. The leucine zipper structure has been found in DNA binding proteins and is thought to facilitate dimerization (15, 18). As these DNA polymerases are monomeric molecules, the leucine zipper structure may have another function for DNA-related enzymes.

The crude enzyme solution prepared as described under Materials and Methods from *E. coli* JM109 cells harboring pUPL1 exhibited very low DNA polymerase activity (about 40 units/ml culture). To increase the expression of the Tth Pol I gene, the upstream region was shortened from 700 bp to 100 bp by exonuclease digestion, resulting in the plasmid pUPL1253 (see Fig. 1). By using this plasmid, the production level was increased to about 300 units/ml culture. To further increase the expression level of the gene, the plasmid pLED-NS (see Fig. 1) was prepared as described under Materials and Methods. *E. coli* cells harboring pLED-NS produced an increased level of heat-stable polymerase activity (about 500 units/ml culture). Figure 3 also shows that the expression level of the pLED-NS system is higher than that of the pUPL1253 system, although the level is still much lower than that of glucose dehydrogenase genes (20) inserted in a similar high-expression vector. It seems that overproduction of Tth Pol I is somewhat harmful to *E. coli*. Nevertheless, compared to the original system of *T. thermophilus* (unpublished results), this expression system has advantages in the high expression level (about 10 times larger). In addition, Fig. 3 shows that heat treatment effectively facilitates the purification of thermostable Pol I by removing other heat-labile proteins in the crude cell extract. We believe that this system will be a good source of enzyme for the study and application of Tth Pol I.

## REFERENCES

1. **Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B., and Erlich, H. A.:** Primer-direct-

ed enzymatic amplification of DNA with a thermostable DNA polymerase. Science, **239**, 487–491 (1988).
2. **Chien, A., Edgar, D. B., and Trela, J. M.:** Deoxyribonucleic acid polymerase from the extreme thermophile *Thermus aquaticus*. J. Bacteriol., **127**, 1550–1557 (1976).
3. **Lawyer, F. C., Stoffel, S, Saiki, R. K., Myambo, K., Drummond, R., and Gelfand, D. H.:** Isolation, characterization, and expression in *Escherichia coli* of the DNA polymerase gene from *Thermus aquaticus*. J. Biol. Chem., **264**, 6427–6437 (1989).
4. **Oshima, T. and Imahori, K.:** Description of *Thermus thermophilus* (Yoshida and Oshima) comb. nov., a nonsporulating thermophilic bacterium from a Japanese thermal spa. Int. J. Syst. Bacteriol., **24**, 102–112 (1974).
5. **Yanisch-Perron, C., Vieria, J., and Messing, J.:** Improved M13 phage cloning and host strains: nucleotide sequences of the M13 mp18 and pUC19 vectors. Gene, **33**, 103–119 (1985).
6. **Miki, T., Yasukochi, T., Nagatani, H., Furuno, M., Orita, T., Yamada, H., Imoto, T., and Horiuchi, T.:** Construction of a plasmid vector for the regulatable high level expression of eukariotic genes in *Escherichia coli*: an application to overproduction of chicken lysozyme. Protein Eng., **1**, 327–332 (1987).
7. **Norrander, J., Kempe, T., and Messing, J.:** Construction of improved M13 vector using oligodeoxynucleotide-directed mutagenesis. Gene, **26**, 101–106 (1983).
8. **Maniatis, T., Fritsch, E. F., and Sambrook, J.:** Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York (1982).
9. **Saito, H. and Miura, K.:** Preparation of transforming deoxynucleic acid by phenol treatment. Biochim. Biophys. Acta, **72**, 619–629 (1963).
10. **Vogelstein, B. and Gillespie, D.:** Preparation and analytical purification of DNA from agarose. Proc. Natl. Acad. Sci. USA, **76**, 615–619 (1979).
11. **Messing, J., Crea, R., and Seeburg, P. H.:** A system for shotgun DNA sequencing. Nucleic Acids Res., **9**, 309–321 (1981).
12. **Ito, J. and Braithwaite, D. K.:** Compilation and alignment of DNA polymerase sequences. Nucleic Acids Res., **19**, 4045–4057 (1991).
13. **Nakahigashi, K. and Inokuchi, H.:** Nucleotide sequence between the *fadB* gene and the *rrnA* operon from *Escherichia coli*. Nucleic Acids Res., **18**, 6439 (1990).
14. **Pearson, W. R. and Lipman, D. J.:** Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. USA, **85**, 2444–2448 (1988).
15. **Landschulz, W. H., Johnson, P. F., and McKnight, S. L.:** The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. Science, **240**, 1759–1764 (1988).
16. **Joyce, C. M., Kelley, W. S., and Grindley, N. D. F.:** Nucleotide sequence of the *Escherichia coli polA* gene and primary structure of DNA polymerase I. J. Biol. Chem., **257**, 1958–1964 (1982).
17. **Ollis, D. L., Brick, P., Hamlin, R., Xuong, N. G., and Steiz, T. A.:** Structure of large fragment of *Escherichia coli* DNA polymerase I complexed with dTMP. Nature, **313**, 762–766 (1985).
18. **Busch, S. J. and Sassone-Corsi, P.:** Dimers, leucine zippers and DNA-binding domains. Trends Genet., **6**, 36–40 (1990).
19. **Laemmli, U. K.:** Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature, **227**, 680–685 (1970).
20. **Mitamura, T., Urabe, I., and Okada, H.:** Enzymatic properties of isozymes and variants of glucose dehydrogenase from *Bacillus megaterium*. Eur. J. Biochem., **186**, 389–393 (1989).