Nucleic Acids Research, 1993, Vol. 21, No. 2 259–265

Organization and nucleotide sequence of the DNA polymerase gene from the archaeon *Pyrococcus furiosus*

Takashi Uemori, Yoshizumi Ishino*, Hiroyuki Toh[1], Kiyozo Asada and Ikunoshin Kato
Biotechnology Research Laboratories, Takara Shuzo Co., Ltd, Otsu, Shiga 520-21 and [1]5th Department, Protein Engineering Research Institute, Suita, Osaka 565, Japan

Received October 1, 1992; Revised and Accepted December 9, 1992 DDBJ accession no. D12983

ABSTRACT

We cloned the gene encoding the thermostable DNA polymerase from the archaeon *Pyrococcus furiosus*. The DNA fragment of 2785 base pair (bp) containing the structural gene for DNA polymerase was sequenced. DNA polymerase (*Pfu* polymerase), as deduced from the DNA sequence, consisted of 775 amino acids, had a molecular weight of 90,109, and was structurally homologous to the α-like DNA polymerases (family B) represented by human DNA polymerase α and *Escherichia coli* DNA polymerase II. An unrooted phylogenetic tree of the α-like DNA polymerases based on the amino acid sequence alignment was constructed. *Pfu* polymerase, with two other archaeon polymerases, constitutes a group with some animal viruses. The transcription initiation sites of the *pol* gene were identified by analysis of *in vivo* transcripts of both from *P.furiosus* and *E.coli*, and the promoters were assigned upstream of the *pol* coding region. A typical promoter sequence for the archaeon was found at a reasonable distance from the transcription initiation site in *P.furiosus*.

INTRODUCTION

DNA polymerase genes from many organisms have been cloned and their deduced amino acid sequences have been compared. On the basis of similarities in these amino acid sequences, DNA polymerases have been classified into two major groups: the *Escherichia coli* DNA polymerase I (Pol I) family and the eukaryotic DNA polymerase α family. A classification of DNA polymerases into families A, B, and C according to the homology of the amino acid sequence with *E.coli* Pol I, II, and III, respectively, has been proposed (1).

Extremely thermostable DNA polymerases have been purified from some archaeon (2–6) and the genes have been cloned (7–9). The deduced amino acid sequences of the DNA polymerases showed that they all belong to the α family (family B). In this study, we report the entire nucleotide sequence of the *P.furiosus* DNA polymerase gene and the deduced primary structure of its protein. We identified the transcription initiation site in both *P.furiosus* and *E.coli*, and located the promoter of the gene.

MATERIALS AND METHODS

Bacterial strain

Pyrococcus furiosus strain Vc1, DSM3638[T] (10) was obtained from the Deutsche Sammlung von Mikroorganismen (Braunschweig, Germany). The strain was grown at 95°C without being shaken, in a broth containing 10 g of tryptone, 5 g of yeast extract, 35 g of the powder conpornent and 5 ml of the solution conpornent of the artificial sea water (Jamarin S, Jamarin Laboratory, Osaka, Japan). , 10 g of starch, and 10 ml of trace elements per 1000 ml. Trace elements (100×conc) contains 1.5 g of nitriloacetic acid, 3.0 g of $MgSO_4$, 1.0 g of NaCl, 0.1 g of $FeSO_4$ $7H_2O$, 0.1 g of $CoSO_4$, 0.1 g of $CaCl_2$ $2H_2O$, 0.1 g of $ZnSO_4$, 10 mg of $CuSO_4$ $5H_2O$, 10 mg of $KAl(SO_4)_2$, 10 mg of H_3BO_3, 10 mg of Na_2MoO_4 $2H_2O$, and 25 mg of $NiCl_2$ $6H_2O$. Nitriloacetic acid was dissolved in KOH to pH 6.5, the minerals were added, the pH was adjusted to 7.0 with KOH, and then the volume was brought to 1000 ml.

Recombinant DNA techniques

DNA was manipulated *in vitro* by standard procedures (11). Restriction enzymes, DNA ligase, reverse transcriptase, the plasmids pUC18, 19, and pTV118N, the universal primers for pUC and pTV vectors, cassettes and cassette primers for *Eco*RI, *Hin*dIII, *Sau*3AI and *Xba*I, and the specific primers for the *P.furiosus pol* gene were products of Takara Shuzo (Kyoto, Japan). [α^{32}P]dCTP and [methyl-^{3}H]TTP were purchased from Amersham International plc (Bucks, UK).

PCR conditions

P. furiosus DNA (0.5 ng) and a cassette oligonucleotide (50 ng) were ligated in 20 μl of reaction solution, then 1 μl of the reaction mix was used as the template for amplification by the PCR. S1 (100 pmol) and cassette primers (25 pmol) were added and 30 cycles were performed with a temperature profile of 30 sec at 94°C, 2 min at 45°C, and 2 min at 72°C in a DNA thermal cycler (Perkin-Elmer Cetus, Norwalk, CT). One microliter of the PCR mixture was used as the template and 25 more cycles were done under the same conditions.

* To whom correspondence should be addressed

DNA sequencing

The nucleotide sequences of the DNA fragments were analyzed by dideoxy nucleotide chain termination method basically as described by Sanger *et al.* (12). Ladderman (*Bca*BEST) DNA polymerase (13) was used for DNA chain elongation reaction.

Amino acid sequence analysis

Partially purified protein was blotted from an SDS-polyacrylamide gel onto a polyvinylidene difluoride membrane as described earlier (14) and the N-terminals were analyzed by Edman degradation with an automated amino acid sequencer (470A, Applied Biosystems, Foster City, CA). The purification procedure of the protein will be described elsewhere.

Analysis of transcripts

The transcription initiation sites of the *pol* gene were identified by primer extension of the in vivo transcripts done basically as described by Treisman *et al.* (15) with the synthetic oligonucleotide dGCCTAATAACAGGTTTTCCT, which is the complementary sequence of the N-terminal coding region of the *pol* gene, as the primer.

Computer analysis

Amino acid sequences, which have been identified as the members of the α-like DNA polymerase family, were collected. Pairwise comparison of the sequences was carried out by a computer program with local alignment algorithm (16). Then, multiple alignments of conserved regions were constructed according to the results of the pairwaise comparison. An unrooted phylogenetic tree was constructed by neighbor-joining method (17) based on differences in amino acid sequences between aligned pairs of α-like DNA polymerases. The differences were calculated from the number of substitutions and the number of aligned sites, which were summed up over the entire aligned segments. In the calculation of the differences, a continuous gap was treated as a single substitution, regardless of its length. The differences were corrected as described by Hood *et. al.* (18). The DNA polymerase sequences compared were the following: *Thermococcus litoralis* (9), *Sulfolobus solfataricus* (8), *Saccharomyces cerevisiae* DNA polymerase II (19). *Autographa californica* nuclear polyhedrosis virus (20), fowlpox virus (21), vaccinia virus (22), chlorella virus (23), REV3, a putative DNA polymerase of *S.cerevisiae* required for inducible mutagenesis (24), Plasmodium falciparum DNA polymerase δ (25), bovine DNA polymerase δ (26), human DNA polymerase δ (27), *Schizosaccharomyces pombe* DNA polymerase δ (28), *S.cerevisiae* DNA polymerase III (29), DNA polymerase from human cytomegalovirus (30), human herpesvirus 6 (31), Epstein-Barr virus (32), varicella-zoster virus (33), herpes simplex virus type 1 (34), herpes simplex virus type 2 (35), *Trypanosoma brucei* (36), *Drosophila melanogaster* DNA polymerase α (37), human DNA polymerase α (38), *S.cerevisiae* DNA polymerase I (39), *S. pombe* DNA polymerase a (40), *E.coli* DNA polymerase II (41), DNA polymerase from *E.coli* phage T4 (42), adenovirus type 12 (43), adenovirus type 5 (44), adenovirus type 2 (45), adenovirus type 7 (46), pClK1, a linear plasmid of the fungus *Claviceps purpurea* (47), pGKL1, a linear plasmid of the yeast *Kluyveromyces lactis* (48), pGKL2 (49), bacteriophage PRD1 (50), φ29 (51), M2 (52), S1 mitochondrial DNA of maize (53), and *Ascobolus immersus* mitochondrion plasmid pAI2 (54).

RESULTS AND DISCUSSION

Isolation of the *pol* gene from *P.furiosus* DNA

Based on the amino-terminal sequence of the purified polymerase protein, the mixed oligonucleotide S1, ATGAT(T/C/A)(T/C)-T(T/C/A/G)GA(T/C)GT(T/C/A/G)GA(T/C)TA, and S2, GA(T/C)TA(T/C)AT(T/C/A)AT(T/C/A/G)GA(A/G)GA (Fig. 1), were synthesized. PCR was performed with primer S1 coupled with one of the cassettes primer C1 (55) from a cassete-ligation-mediated library of *P.furiosus* DNA as the template. Specific fragments 970 and 600 bp long were amplified from the *Eco*RI and the *Bam*HI cassette libraries, respectively. With the 970-bp fragment as the probe, a *P.furiosus* genomic library consisting of a cosmid vector (Triple helix cosmid vector, Stratagene Cloning Systems, La Jolla, CA) that had fragments with the mean size of 40 kbp from *Sau*3AI (Asada *et al.*, unpublished) was screened by colony hybridization, and a positive clone that contained a 40-kbp fragment of *P.furiosus* was obtained. To locate of the *pol* structural gene, the positive cosmid DNA was digested with *Eco*RI, *Bam*HI, *Hin*dIII, and *Pst*I and

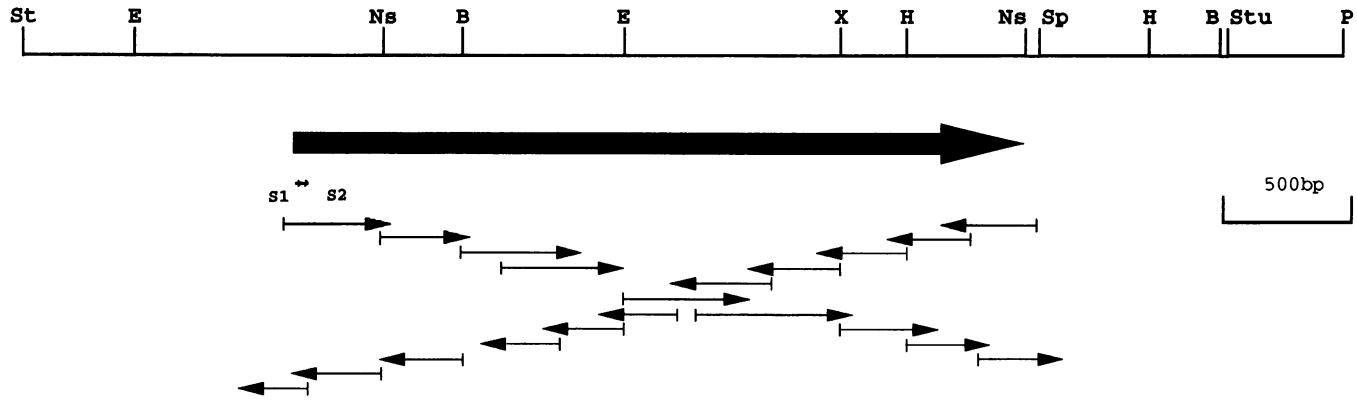


Figure 1. Physical map of a DNA fragment containing the *Pfu pol* gene, and DNA sequencing strategy. Arrows indicate the extent of the nucleotide sequences determined. The thick arrow indicates the coding region and orientation of the *pol* gene. Abbreviations: B, *Bam*HI; E, *Eco*RI; H, *Hin*dIII; Ns, *Nsp*7524V; P, *Pst*I; Sp, *Sph*I; St, *Stu*I; X, *Xba*I.

analyzed by Southern hybridization. A positive *Pst*I fragment of 7.0 kbp, was isolated and inserted into the *Pst*I site of the plasmid pTV118N. The isolated plasmid, designated pPF100, was used as a template to confirm that specific fragments could be amplified by PCR with the primer S1 and the universal primer M4 from the vector plasmid. PCR was also helpful in prediction of the region of the *pol* structural gene in the inserted fragment of 7.0 kbp. The extract from *E.coli* strains carrying the plasmid pPF100 had DNA polymerase activity at 75°C, and therefore we concluded that pPF100 contained the *pol* gene of *P.furiosus*.

Nucleotide sequence of the *pol* gene

Because restriction enzyme cutting sites were scattered in the 7.0-kbp fragment (Fig. 1), the restriction fragments *Bam*HI-*Bam*HI, *Bam*HI-*Eco*RI, *Bam*HI-*Hin*dIII, *Eco*RI-*Eco*RI, *Eco*RI-*Hin*dIII, *Eco*RI-*Sph*I, *Hin*dIII-*Sph*I, *Nco*I-*Nco*I, *Nsp*I-*Stu*I, *Sph*I-*Xba*I, and *Xba*I-*Xba*I were excised and subcloned into pUC118 and 119. Dideoxy sequencing was performed with universal primers M4, and RV for the plasmid vectors. To check the sequence, specific primers were synthesized on the basis of the sequences that were already read, and finally both strands were

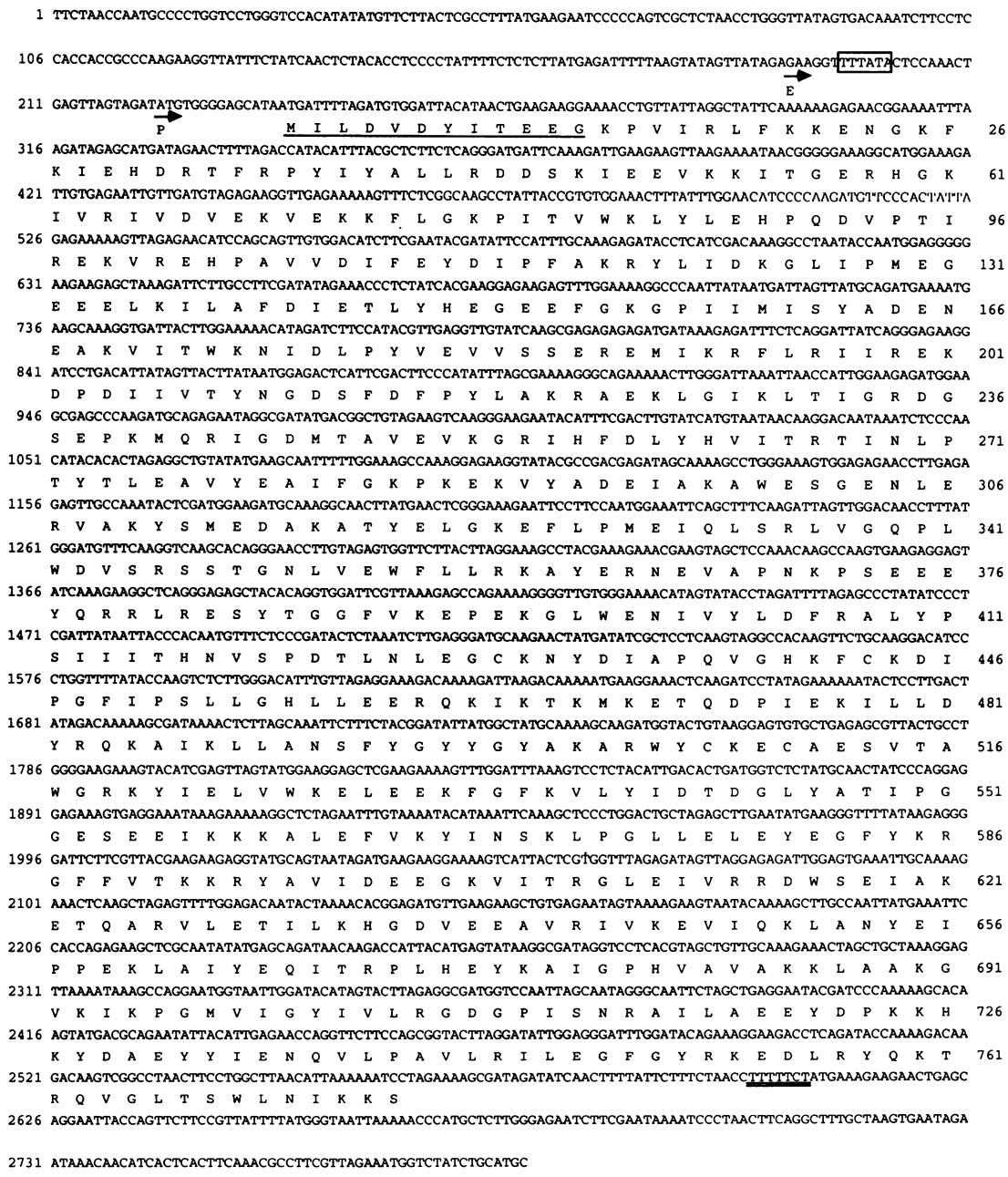


Figure 2. Nucleotide sequence of the *pol* gene and the deduced amino acid sequence of the DNA polymerase. The transcriptional initiation sites in *P.furiosus* and *E.coli* are indicated by arrowheads labelled P and E, respectively. The sequences corresponding to the promoter (box A) are boxed and the sequence corresponding to the terminator is doubly underlined. The amino-terminal amino acid sequences identified are underlined.

read entirely and 2825 bases were determined (Fig. 2). Possible translational regions were searched for and one large open reading frame, the deduced amino-terminal sequence of which exactly matched that obtained by protein analysis, was found within the sequenced region. Therefore, ATG at 237−239 was determined to be the initiation codon. The encoded protein consists of 775 amino acids and the calculated molecular weight of that was 90,109, which agreed with the apparent molecular weight estimated by SDS-polyacrylamide gel electrophoresis (PAGE). The GC contents of the entire sequenced region and the *pol* structural gene were 40.0 and 38.5 mol%, respectively, compared with 37.1 mol% for genomic *P.furiosus* Vc1 (10). One feature of interest was that the coding strand contained many adenine residues (37.4 mol%), which caused some unbalanced codon usage, for example, codons for arginine with C as the first letter were very rare. AGA and AGG were abundant as in other archaeal genes (56).

Identification of the transcription initiation site

The transcription initiation site of the *pol* gene was determined by primer extension to identify the promoter of the gene. Synthetic oligonucleotides with a sequence coresponding to that of the N-terminal coding region of the *pol* gene were labeled with ^{32}P and hybridized with the total RNA extracted from *P.furiosus* and *E.coli* carrying pPF100. The size of the DNA after extention with reverse transcriptase was estimated by comparison with the dideoxy sequence ladders of the corresponding region analyzed with the same labeled primers. Adenine at 223 and guanine at 190 were the transcription initiation sites of the *pol* gene in *P.furiosus* and *E.coli*, respectively (Fig. 3). From these results, promoter sequences of the *pol* gene were assigned. The typical sequence of the archaeal promoter, TTTATA (box A) (57,58), was 26 bases upstream of the transcription initiation site in *P.furiosus*, and a box-B-like sequence was also found around the initiation site. However, in *E.coli*, TTCTCT (155−160) and TATAGT (178−183) were preferentially used as the −35 and −10 regions of the bacterial promoter. Downstream of the termination codon, TTTTTCT (2601−7), a perfect match of a proposed terminator sequence in archaea (59), was found.

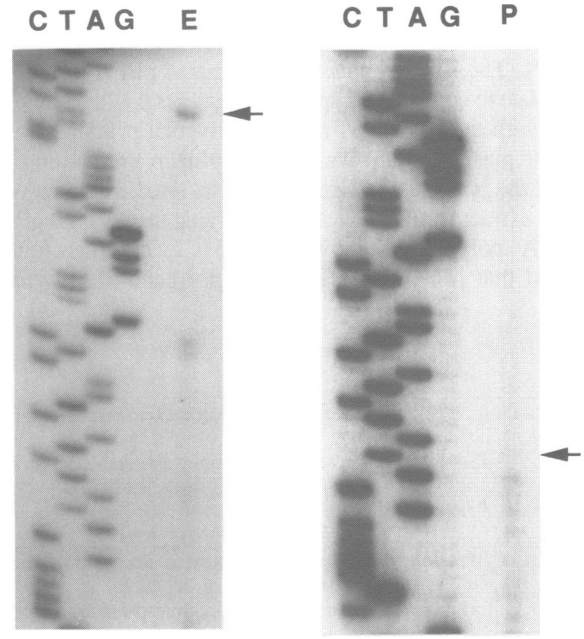
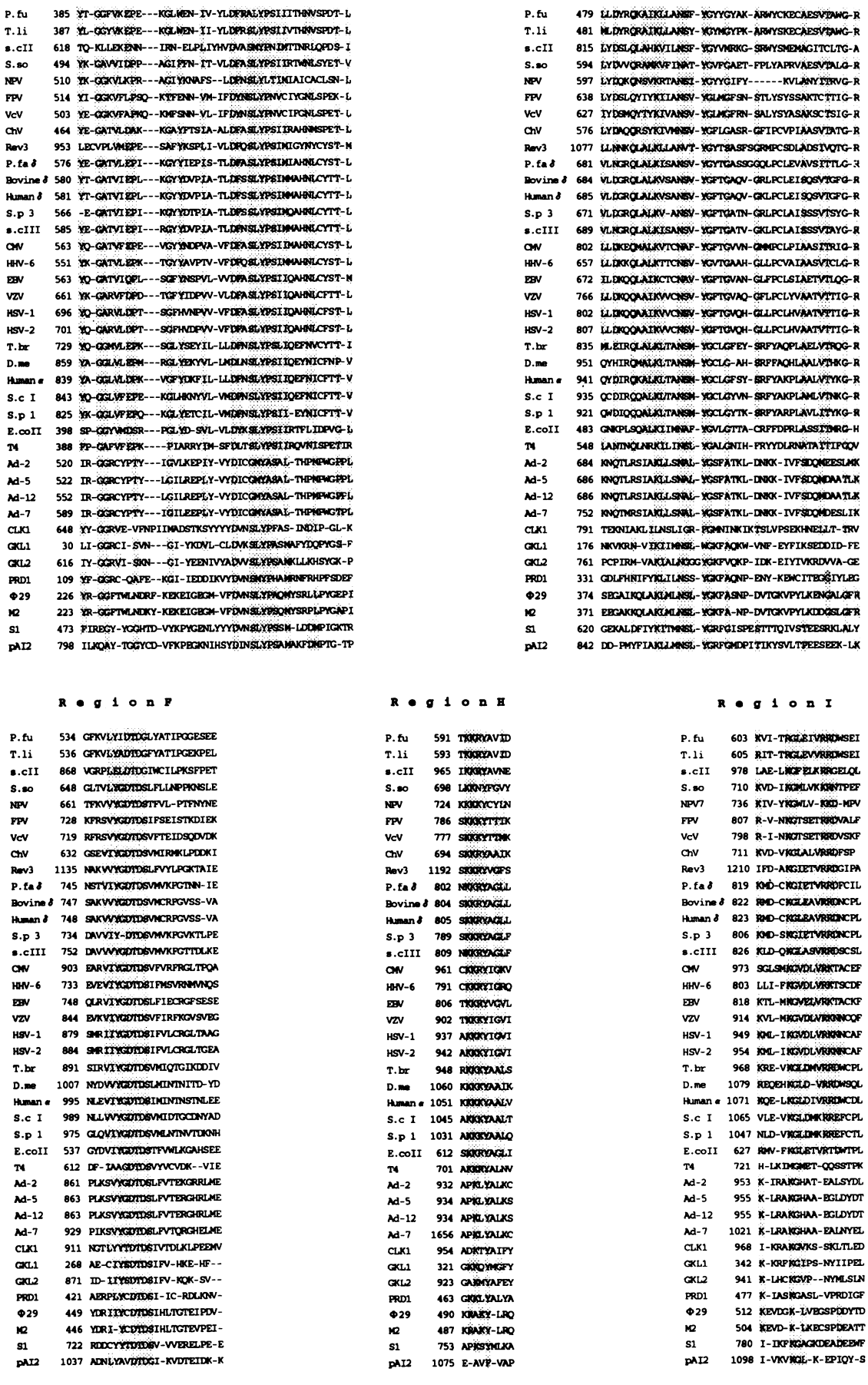


Figure 3. Analysis of the transcription initiation site of the *pol* gene by the primer extension method. Lanes G, A, T, and C show ladders of the dideoxy sequencing reaction. The products of primer extension on the mRNA extracted from *P.furiosus* cells (lane P) and from *E.coli* (lane E) were analyzed by electrophoresis on a polyacrylamide gel.

Phylogenetic relationship of *P.furiosus* DNA polymerase with the α-like (family B) DNA polymerase

We found that *E.coli* DNA polymerase II belongs to the family of α-like DNA polymerases, by a computer-assisted homology search (41). From its amino acid sequence and the aphidicolin sensitivity of its polymerizing activity, *Pfu* polymerase seems to be included in this family. Therefore, the amino acid sequences of all of the α-like DNA polymerases that are available so far were aligned by the method described earlier (41), and regions C, E, F, and I were found to be conserved, as they are in all α-like DNA polymerases so far reported (Fig. 4). On the basis of this alignment, we again drew an unrooted phylogenetic tree of α-like DNA polymerases (Fig. 5). All three archaeal DNA polymerases including *pfu* polymerase, constituted a group with some animal viruses and this group was different from groups of eukaryotic Pols α and δ. *Pyrococcus* and *Thermococcus* are extreme thermophiles and their DNA polymerases were 74.5% identical when a continuous gap was calculated as one substitution at one site (data not shown), although there were large differences in the genes. *P.furiosus* does not have any intervening sequences (IVS) in the *pol* gene, in contrast to the gene of *T.litoralis* (9). The *pol* gene of *T.litoralis* carries two IVS, which appears to be transcribed and translated but then removed by protein splicing. *Pyrococcus* and *Themococcus* have similar morphology, however their optimum temperatures and the extent of their tolerance to ionic strength while growing are different, as is the GC content of their genomic DNA (10). *Sulfolobus* is also a

Figure 4. Regions with high similarity in *Pfu* polymerase and with other α-like DNA polymerases (family B). Amino acids with similar properties are grouped as follows: LIMV, SPTAG, YWF, DEQN, and KRH. Amino acid positions that are occupied more than 70% with similar amino acids among the top 27 sequences are shadowed. Abbreviations: P. fu, *Pfu* polymerase; T. li, *T. litoraris*, S. c, *S.cerevisiae* Pol II; S. so, *S.solfataricus*; NPV, *Autographa californica* nuclear polyhedrosis virus; PFV, fowlpox virus; VcV, vaccinia virus; ChV, chlorella virus; REV3, putative DNA polymerase required for inducible mutagenesis of *S.cerevisiae;* P. fa. δ, *Plasmodium falciparum* Pol δ; bovine δ, calf thymus Pol δ; human δ, human Pol δ; S. p 3, *Schizosaccharomyces pombe* Pol δ; S. c III, *S.cerevisiae* Pol III; CMV, human cytomegalovirus; HHV6, human herpes virus 6; EBV, Epstein-Barr virus; VZV, varicella-zoster virus; HSV1, herpes simplex virus type 1; HSV2, herpes simplex virus type 2; T. br, *Trypanosoma brucei*; D. me, *Drosophila melanogaster* Pol α; human α, human Pol α; S. c I, *S.cerevisiae* Pol I; S. p 1, *S. pombe* Pol 1; E. co II, *E.coli* Pol II; T4, *E.coli* phage T4; Ad, adenovirus (types 12, 5, 2, and 7); CKL1, a linear plasmid of the fungus *Claviceps purpurea*; GKL, a linear plasmid of the yeast *Kluyveromyces lactis* (1 and 2); PRD1, bacteriophage PRD1; φ29, bacteriophage φ29; M2, bacteriophage M2; S1, S1 mitochondrial DNA of maize; PAI2, *Ascobolus immersus* mitochondrion plasmid pAI2.


Region C
Region E
Region F
Region H
Region I

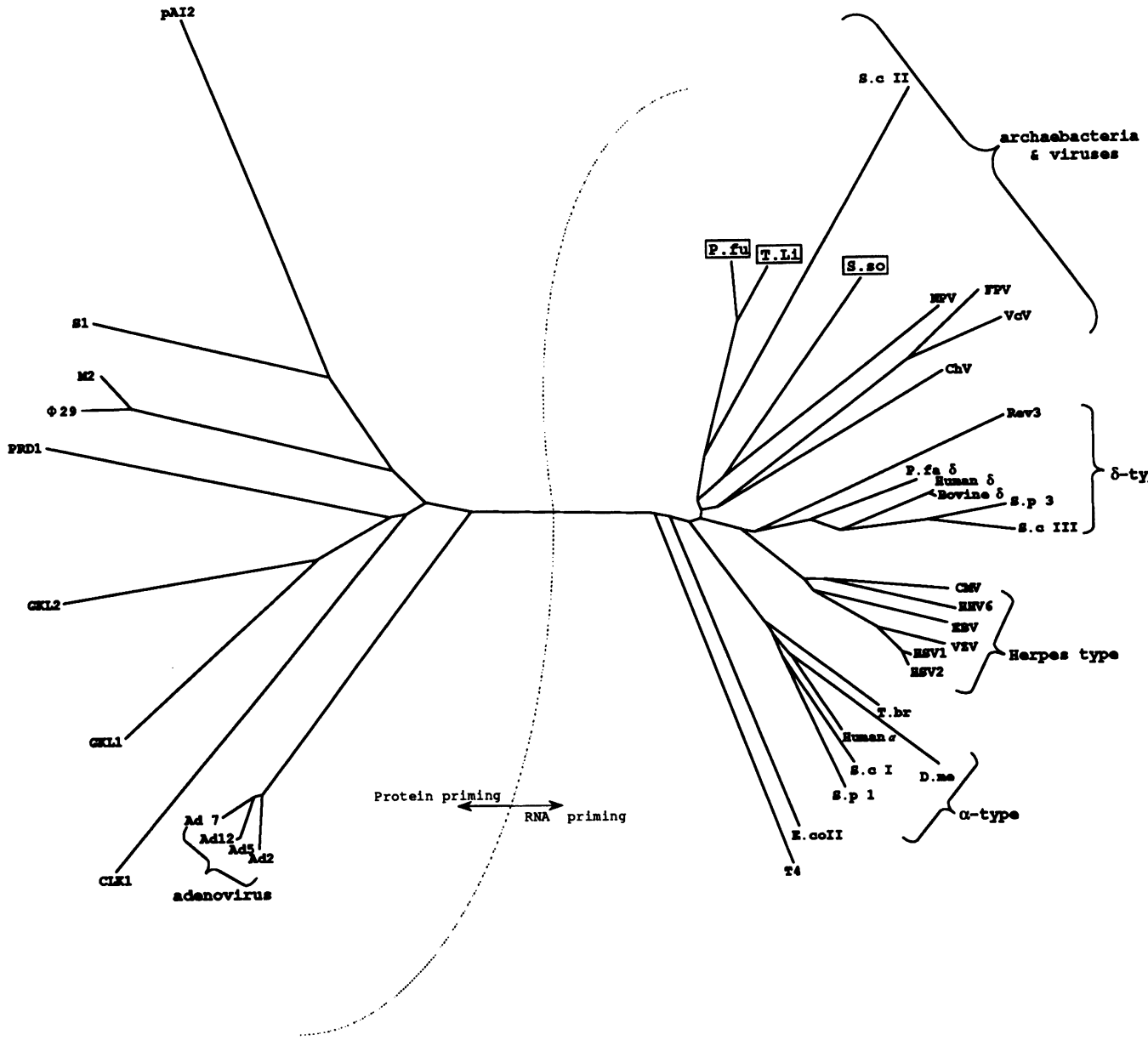


Figure 5. Unrooted phylogenetic tree of α-like DNA polymerase. Calculations for the construction of the phylogenetic tree were as described in Materials and Methods. Abbreviations are the same as those in Fig. 4.

thermophilic archaeon, with a slightly lower temperature of optimal growth (80°C) than the other two archaeon. The *pol* gene of that organism also lacks an IVS although the deduced amino acid sequence is less similar to that of *Pfu* gene than that of the *Tli*. It would be interesting to study the structure of the genomic DNA of these bacteria by pulsed field gel electrophoresis. It is not known why *S.cerevisiae* Pol II is so similar to archaeal polymerases. The α-like DNA polymerases (family B) can be divided into two subfamilies, one of polymerases that are protein-primed and the other of polymerases that are RNA-primed (41); we identified four groups, archeon and viruses, Pol δ, Herpes, and Pol α, in the RNA primed subfamily in this study.

Sequence alignment of all of the members of the α-like DNA polymerases that we studied did not make possible the identification of the essential region for the 3′→5′ exonuclease activity, because some of the polymerases do not have that activity in the same polypeptide. Morrison *et al.* (60) proposed a motif, Phe-Asp-Ile-Glu-Thr, as a part of the 3′→5′ exonuclease active site of α-like DNA polymerases, and proved that the Asp and Glu were critical residues for the exonuclease activity in *S.cerevisiae* Pol II by site-directed mutagenesis. Simon *et al.* also found the same results for *S.cerevisiae* Pol III (61). We found this motif in the sequence of *E.coli* Pol II and confirmed that the Asp and Glu in the motif were essential for the exonuclease activity by site-directed mutagenesis (Ishino *et al.*, unpublished). This motif was also found in the *Pfu* polymerase at 140 to 144. The *pfu* polymerase has associated 3′→5′ exonuclease activity (6), so this region must be part of the exonuclease active site.

The α-like DNA polymerases have not been studied by crystallography. In this study, we succeeded to overproduce *Pfu* polymerase in *E.coli*, which may be helpful for structural analysis. It would be of use for the understanding of its thermophilisity to analyze the structure of this protein.

ACKNOWLEDGEMENT

We thank Kazuko Nakanishi for preparing the figures.

REFERENCES

1. Ito, J. and Braithwaite, D. K. (1991) *Nucleic Acids Res.*, **19**, 4045–4057.
2. Elie, C., De Recondo, A. M., and Forferre, P. (1989) *Eur. J. Biochem.*, **178**, 619–626.
3. Klimczak, L. J., Grummt, F. and Burger, K. J. (1985) *Nucleic Acids Res.*, **13**,5269–5282.
4. Rossi, M., Rella, R., Pensa, M., Bartolucci, S., De Rosa, M., Gambacorta, A.,Raia, C. A., and Dell'. Aversano Orabona, N. (1986) *System . Appl. Microbiol.*, **7**, 337–341.
5. Rella, R., Raia, C. A., Pisani, F. M., D'Auria, S., Nucci, R., Gambacorta, A., De Rosa, M. and Rossi, M. (1990) *Int. J. Biochem.* **39**, 83–99.
6. Lundberg, K. S., Shoemaker, D. D., Adams, M. W. W., Short, J. M., Sorge, J. A. and Mathur, E. J. (1991) *Gene*, **108**, 1–6.
7. Mathur, E. J., Adams, M. W. W., Callen, W. N. and Cline, J. (1991) *Nucleic Acids Res.*, **19**, 6952.
8. Pisani, F. M., De Martino, C. and Rossi, M. (1992) *Nucleic Acids Res.*, **20**, 2711–2716.
9. Perler, F. B., Comb, D. G., Jack, W. E., Moran, L. S., Qiang, B., Kucera, R. B., Benner, J., Slatko, B. E., Nwankwo, D. O., Hempstead, S. K., Carlow, C. K. S. and Jannasch, H. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 5577–5581.
10. Fiala, G. and Stetter, K. O. (1986) *Arch. Microbiol.*, **145**, 56–61.
11. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) in *Molecular cloning: a laboratory manual.* Cold Spring Harbor Laboratory, Cold Spring Harbor, N. Y.
12. Sanger, F., Nicklen, S. and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
13. Ishino, Y. (1992) *Amer Biotechnol Lab.*, **10**, 47.
14. Matsudaira, P. (1987) *J. Biol. Chem.*, **262**, 10035–10038.
15. Treisman, R., Proudfoot, N. J., Shander, M. and Maniatis, T. (1982) *Cell*, **29**, 903–911.
16. Nishikawa, K., Nakashima, H., Kanehisa, M. and Ooi, T. (1987) *Protein Seq. DataAnal.*, **1**, 107–116.
17. Saitou, N. and Nei, M. (1987) *Mol. Biol. Evol .*, **4**, 406–425.
18. Hood, L. E., Wilson, J. H., and Wood, W. E. (1975) *Molecular biology of eucaryotic cells: a problems approach.* W. A. Benjamin, Menlo Park, California.
19. Morrison, A., Araki, H., Clark, A. B., Hamatake, R. K. and Sugino, A. (1990) *Cell*, **62**, 1143–1151.
20. Tomalski, M. D., Wu, J. and Miller, L. K. (1988) *Virology*, **167**, 591–600.
21. Binns, M. M., Stenzler, L., Tomley, F. M., Campbell, J. and Boursnell, M. E. G. (1987) *Nucleic Acids Res.*, **15**, 6563–6573.
22. Earl, P. L., Jones, E. V. and Moss, B. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 3659–3663.
23. Grabherr, R., Strasser, P. and Van Etten, J. L. (1992) *Virology*, **188**, 721–731.
24. Morrison, A., Christensen, R. B., Alley, J., Beck, A. K., Bernstine, E. G., Lemontt, J. F., and Lawrence, C. W. (1989) *J. Bacteriol.*, **171**, 5659–5667.
25. Fox, B. A. and Bzik, D. J. (1991) *Mol. Biochem. Parasitol .*,49, 289–296.
26. Zhang, J., Chung, D. W., Jan, C-K., Downey, K. M. Davie, E. W. and So, A. G. (1991) *Biochemistry*, **30**, 11742–11750.
27. Yang, C-L., Chang, L-S., Zhang, P., Hao, H., Zhu, L., Toomey, N. L. and Lee, M. Y. W. T. (1992) *Nucleic Acids Res.*, **20**, 735–745.
28. Pignède, G., Bouvier, D., de Recondo, A-M. and Baldacci, G. (1991) *J. Mol. Biol.*, **222**, 209–218.
29. Boulet, A., Simon, M, Faye, G., Bauer, G. A. and Burgers, P. M. J. (1989) *EMBO J.*, **8**, 1849–1854.
30. Kouzarides, T., Bankier, A. T., Satchwell, S. C., Weston, K, Tomlinson, P. anBarrell, B. G. (1987) *J. Virol.*, **61**, 125–133.
31. Teo, I. A., Griffin, B. E. and Jones, M. D. (1991) *J. Virol.*, **65**, 4670–4680.
32. Baer, R., Bankier, A. T., Biggin, M. D., Deininger, P. L., Farrell, P. J., Gibson, T. J., Hatfull, G., Hudson, G. S., Satchwell, S. C., Sèguin, C., Tuffnell, P. S. and Barrell, B. G. (1984) *Nature*, 310, 207–211.
33. Davison, A. J. and Scott, J. E. (1986) *J. Gen. Virol.*, **67**, 1759–1816.
34. Larder, B. A., Kemp, S. D. and Darby, G. (1987) *EMBO J.*, **6**, 169–175.
35. Tsurumi, T., Maeno, K. and Nishiyama, Y. (1987) *Gene*, **52**, 129–137.
36. Leegwater, P. A. J., Strating, M. S., Murphy, N. B., Kooy, R. F., van der Vliet, P. C.and Overdulve, J. P. (1991) EMBL Data Library
37. Hirose, F., Yamaguchi, M., Nishida, Y., Masutani, M., Miyazawa F., Hanaoka, F. and Matsukage, A. (1991) *Nucleic Acids Res.*, **19**, 4991–4998.
38. Wong, S. W., Wahl, A. F., Yuan, P-M., Arai, N., Pearson, B. E. Arai, K., Korn, D., Hunkapiller, M. W. and Wang, T. S. F. (1988) *EMBO J.*, **7**, 37–47.
39. Pizzagalli, A., Valsasnini, P., Plevani, P. and Lucchini, G. (1988) *Proc. Natl. Acad. Sci. USA.*, **85**, 3772–3776.
40. Damagnez, V., Tillit, J., de Recondo, A-M. and Baldacci, G. (1991) *Mol. Gen. Genet.*, **226**, 182–189.
41. Iwasaki, H., Ishino, Y., Toh, H., Nakata, A. and Shinagawa, H. (1991) *Mol. Gen. Genet.*, **226**, 24–33.
42. Spicer, E. K., Rush, J., Fung, C., Reha-Krantz, L. J., Karam, J. D. and Konigsberg, W. H. (1988) *J. Biol. Chem.*, **263**, 7478–7486.
43. Shu, L., Hong, J. S., Wei, Y-f. and Engler, J. A. (1986) *Gene*, **46**, 187–195.
44. Dekker, B. M. M. and van Ormondt, H. (1984) *Gene*, **27**, 115–120.
45. Gingeras, T. R., Sciaky, D., Gelinas, R. E., Bing-Dong, J., Yen, C. E., Kelly, M. M.,Bullock, P. A., Parsons, B. L., O'Neill, K. E. and Roberts, R. J. (1982) *J. Biol. Chem.*, **257**, 13475–13491.
46. Engler, J. A., Hoppe, M. S. and van Bree, M. P. (1983) *Gene*, **21**, 145–159.
47. Oeser, B. and Tudzynski, P. (1989) *Mol. Gen. Genet.* 217, 132–140.
48. Stark, M. J. R., Mileham, A. J., Romanos, M. A. and Boyd, A. (1984) *Nucleic Acids Res.*, **12**, 6011–6030.
49. Tommasino, M., Ricci, S. and Galeotti, C. L. (1988) *Nucleic Acids Res.* **16**, 5863–5878.
50. Jung, G., Leavitt, M. C., Hsieh, J-C. and Ito, J. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 8287–8291.
51. Yoshikawa, H. and Ito, J. (1982) *Gene*, **17**, 322–335.
52. Matsumoto, K., Takano, H., Kim, C. I. and Hirokawa, H. (1989) *Gene*, **84**, 247–255.
53. Pailland, M., Sederoff, R. R. and Levings, C. S., III. (1985) *EMBO J.*, **4**, 1125–1128.
54. Kempken, F., Meinhardt, F. and Esser, K. (1989) *Mol. Gen. Genet.* **218**, 523–530.
55. Isegawa, Y., Sheng, J., Sokawa, Y., Yamanishi, K., Nakagomi, O. and Ueda, S. (1992) *Mol. Cell. Probes*, in press.
56. Cubellis, M. V., Rozzo, C., Montecucchi, P. and Rossi, M. (1990) *Gene*, 89–94.
57. Reiter, W-D., Palm, P. and Zillig, W. (1988) *Nucleic Acids Res.*, **16**, 1–19.
58. Thomm, M. and Wich, G. (1988) *Nucleic Acids Res.*, **16**, 151–163.
59. Reiter, W-D., Palm, P. and Zillig, W. (1988) *Nucleic Acids Res.*, **16**, 2445–2459.
60. Morrison, A., Bell, J. B., Kunkel, T. A. and Sugino, A. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 9473–9477.
61. Simon, M., Giot, L. and Faye, G. (1991) *EMBO J.*, **10**, 2165–2170.