Proc. Natl. Acad. Sci. USA
Vol. 88, pp. 11197–11201, December 1991
Biochemistry

Primary structure of the catalytic subunit of human DNA polymerase δ and chromosomal location of the gene

(cDNA cloning/cDNA sequence/chromosomal location)

DOMINIC W. CHUNG*, JIAN ZHANG†, CHENG-KEAT TAN‡, EARL W. DAVIE*, ANTERO G. SO†‡, AND KATHLEEN M. DOWNEY†‡

*Department of Biochemistry, University of Washington, Seattle, WA 98195; and Departments of ‡Medicine and †Biochemistry/Molecular Biology, University of Miami, Miami, FL 33101

Contributed by Earl W. Davie, September 17, 1991

ABSTRACT **The catalytic subunit ($M_r \approx 124,000$) of human DNA polymerase δ has been cloned by PCR using poly(A)$^+$ RNA from HepG2 cells and primers designed from the amino acid sequence of regions highly conserved between bovine and yeast DNA polymerase δ. The human cDNA was 3443 nucleotides in length and coded for a polypeptide of 1107 amino acids. The enzyme was 94% identical to bovine DNA polymerase δ and contained the numerous highly conserved regions previously observed in the bovine and yeast enzymes. The human enzyme also contained two putative zinc-finger domains in the carboxyl end of the molecule, as well as a putative nuclear localization signal at the amino-terminal end. The gene coding for human DNA polymerase δ was localized to chromosome 19.**

DNA polymerase δ (pol δ) is one of at least two DNA polymerases required to replicate chromosomal DNA in eukaryotic cells (1–4). Similar to many other replication proteins, it is both structurally and functionally conserved among phylogenetically divergent species. The pol δ enzymes from both higher and lower eukaryotes are heterodimers with subunits of ≈125 and 50 kDa (5–9). The 125-kDa subunit contains the polymerase active site and most likely the active site for the 3′–5′ exonuclease activity (10, 11). The latter activity serves as a proofreading function during DNA replication.

The heterodimeric core enzymes from both yeast and bovine sources are inherently nonprocessive but become highly processive in the presence of an accessory protein called the proliferating cell nuclear antigen (PCNA) (12–14). Furthermore, the processivity of yeast pol δ is increased by bovine PCNA and *vice versa* (14, 15), suggesting that the domains involved in protein–protein interactions between pol δ and PCNA must be highly conserved.

Recent isolation of a cDNA coding for the catalytic subunit of bovine pol δ has allowed comparison of the primary structure of this DNA polymerase with that of yeast pol δ, as well as other eukaryotic and prokaryotic DNA polymerases (16). The bovine and yeast polypeptides are 44% identical. Furthermore, the δ polymerases share several regions of sequence similarity with the herpes virus family of DNA polymerases—i.e., those from herpes simplex virus, Epstein–Barr virus, and human cytomegalovirus, in addition to the seven highly conserved regions present in the Tyr-Gly-Asp-Thr-Asp-Ser class of DNA polymerases. These highly conserved regions were originally identified by Wong *et al.* (17) and Spicer *et al.* (18).

To carry out studies on the transcriptional regulation of pol δ, as well as to provide information on structural domains that may be involved in protein–protein interactions between the enzyme and its accessory proteins, the nucleotide sequence of a full-length cDNA coding for the catalytic subunit of human pol δ has been determined.§ Also, the gene coding for this polypeptide has been localized to chromosome 19.

MATERIALS AND METHODS

Poly(A)$^+$ RNA Isolation. Total cellular RNA of human HepG2 cells was from Joost Meijers (University of Washington). Poly(A)$^+$ RNA was selected according to Sambrook *et al.* (19).

Oligonucleotide Synthesis. Oligonucleotides were synthesized by the phosphoramidite method with an Applied Biosystems 380B DNA synthesizer. The oligonucleotides used for DNA amplification were designed with an *Eco*RI or *Bam*HI restriction site at the 5′ end to facilitate subsequent cloning.

DNA Amplification. Poly(A)$^+$ RNA was reverse transcribed by Superscript RNase H$^-$ reverse transcriptase in the presence of the oligo(dT)-containing primer EDT (see Table 1) according to conditions recommended by the supplier (GIBCO). The reaction was stopped by phenol/chloroform extraction, and the cDNA was precipitated by ethanol. The pellet was resuspended in 100 μl of water, and 1 μl was used for the PCR as described (16). DNA fragments amplified by PCR were extracted with phenol/chloroform, precipitated with ethanol, and then digested with the appropriate restriction enzyme; the fragments were fractionated by electrophoresis in a 1% low-melting-point agarose gel. The appropriate DNA bands were excised for subsequent cloning into M13mp18 or -19.

Cloning of Human pol δ cDNA. Overlapping amplified DNA fragments from the first-strand cDNA were cloned into M13mp18 or -19 (Pharmacia LKB). The 5′ and 3′ ends of the human pol δ cDNA were amplified and cloned by a modified rapid amplification of cDNA end (RACE) protocol as described (16). The first-strand cDNA synthesized as described above was purified from the primer and hydrolyzed RNA by chromatography on Sepharose CL-6B. Oligo(dC) extensions were added to the 3′ end of the cDNA by terminal deoxynucleotidyltransferase (GIBCO) in the presence of 1 mM dCTP. The 3′ end of the human pol δ cDNA was amplified with the tailed cDNA using primer HGF2 and an oligo(dT)-containing primer EDT, whereas the 5′ end was amplified using HGR2 and an oligo(dG)-containing primer BAG. Amplified fragments were cloned into M13mp18, as described above. Recombinant plaques containing the authentic 3′ end of the human pol δ cDNA were identified by hybridization to

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: PCNA, proliferating cell nuclear antigen; pol δ, DNA polymerase δ.
§The sequence reported in this paper has been deposited in the GenBank data base (accession no. M80397).

Table 1. Peptide and oligonucleotide sequences used in cloning of human DNA polymerase δ

Bovine peptide sequence	Primer	Sense	Oligonucleotide sequence 5′ → 3′
SLYPSIMMA	H2	Sense	TTTGAATTC-TCG-CTG-TAC-CC(C,G)-TTC-AT(T,C)-ATG-ATG-GC
VVYGDTDSVM	HR1	Antisense	TTTGAATTC-ATG-ACC-GAG-TCT-GTG-TCA-CCA-TAC-ACC-AC
SFDIECAGR	H3	Sense	TTTGAATTC-AGC-TTC-GAC-ATC-GAG-TGC-GC(T,C)-GG(T,C)-(A,C)G
MMAHNLCYTTL	HR3	Antisense	TTTGAATTC-AGT-GTG-GTG-TAG-CAC-AGG-TT(A,G)-TG(A,G)-GCC-ATC-AT
SRLMTQCQRC	HR4	Antisense	TTTGAATTC-GCA-GCG-CTG-GCA-CTG-GGT-CCA-CAG-(C,G)CG-(G,T)GA
SLPIDTQYYLEQQ	J3	Sense	TTTGAATTC-AGC-CTG-CCA-ATC-GAC-ACC-CAG-TAC-TAC-CTG-GA(A,G)-CA(G,A)-CA
	HGF1	Sense	TTTGAATTCGACGAGTTTGTGAAGACCTCAGTG
	HGR1	Antisense	TTTGAATTCTTCTCGTAGCTCTGCACCTTGGC
	HGF2	Sense	TTTGAATTCGCCTTCGCCAAACGCCGCAACTGC
	HGR2	Antisense	TTTGAATTCGAAGCCGTGGATGTGGCAGCAGAC
	HGR3	Antisense	TTTGAATTCACAGCCTCGGCACGGCCCTCGCC
	HGR4	Antisense	TTTGAATTCGCAAGGTCACCAGGCCTCAGGTCC
	HGR5	Antisense	TTTGAATTCAGCAAAAGTCCAGAACTTTATTAA
	DF1	Sense	TTTGAATTCAGTCAGGGGTCACGGCGGCGTGG
	EDT	Antisense	TTTGAATTCTTTTTTTTTTTTTTTTTT
	BAG	Sense	TTTGGATCCGGGGGGGGGGGGGGG

^{32}P-labeled oligonucleotide HR4. After the cDNA sequence of human pol δ was determined from overlapping fragments generated by PCR, two specific primers, DF1 and HGR5, were used in a separate PCR reaction to amplify the entire coding region in a continuous segment. This fragment was first cloned in M13mp18, and the sequence was determined. The complementary strand of the single-stranded M13mp18 recombinant phage that contained the entire coding region was synthesized with the Klenow fragment of DNA polymerase I (GIBCO) by using the universal sequencing primer. The double-stranded insert was excised by *Eco*RI digestion, purified from low-melting-point agarose, and cloned into pUC18 (Pharmacia LKB).

DNA Sequencing and Sequence Analysis. Single-stranded templates of overlapping fragments of the human pol δ cDNA were partially sequenced by the dideoxynucleotide chain-termination method (20) with the universal sequencing primer and specific synthetic oligonucleotide primers. Double-stranded template (insert-containing plasmids) was first denatured with 0.2 M NaOH/0.2 mM EDTA at 37°C for 5 min, neutralized with sodium acetate, pH 5.0, precipitated with ethanol, and then sequenced. All sequencing reactions were done with deoxyadenosine 5′-[α-[^{35}S]thio]triphosphate (Amersham) and engineered T7 DNA polymerase (Sequenase, version 2.0; United States Biochemical). Compression was eliminated by using 7-deazadeoxyguanosine triphosphate and dITP in the sequencing reactions. The DNA sequence and deduced amino acid sequence were analyzed on a microcomputer with the GENEPRO program (Riverside Scientific, Seattle) and Protein Identification Resource protein sequence data base release 26.

Chromosomal Localization. The chromosomal location of the human pol δ gene was determined by PCR analysis of DNA from a panel of 24 human–hamster hybrid cells (Bios, New Haven, CT) as described by Tait *et al.* (21). Oligonucleotide primers HGF2 and HGR4 were used in PCRs at an annealing temperature of 65°C. Karyotype of each cell line was provided by the supplier.

RESULTS AND DISCUSSION

Primary Sequence of the Catalytic Subunit of Human pol δ. A full-length cDNA encoding the catalytic subunit of human pol δ has been isolated by using PCR methodology. Amino acid sequences of regions highly conserved between bovine and yeast pol δ were used to design oligonucleotide primers (H2, H3, HR1, HR4, and HR3) of low degeneracy based on human preferred codon use (Table 1). In addition, primers containing bovine nucleotide sequences (DF1, J3), primers containing authentic human sequences (HGF1, HGF2, HGR1, HGR2, HGR3, HGR4, and HGR5), an oligo(dT)-containing primer EDT, and an oligo(dG)-containing primer BAG (Table 1) were used in PCRs to generate overlapping regions of the human pol δ cDNA, according to the strategy shown in Fig. 1. The entire cDNA sequence, deduced from these overlapping clones, showed that the human cDNA was identical in length in the 5′ noncoding region to that of bovine pol δ. Also, the two cDNAs differed in two nucleotides in this region—i.e., nucleotide 22 contained adenine in the human cDNA and guanine in the bovine cDNA, and nucleotide 39 contained guanine in the human cDNA and adenine in the bovine cDNA. Consequently, the bovine primer DF1 was used successfully with primer HGR5 to amplify the entire coding region of the human pol δ cDNA in one continuous

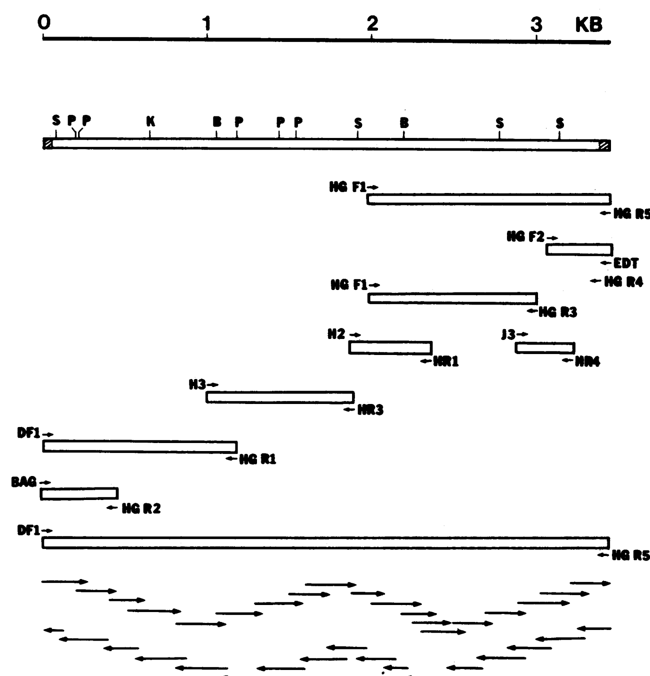

FIG. 1. Partial restriction map and cloning strategy of the cDNA for the catalytic subunit of human pol δ. Overlapping fragments of the cDNA were generated by the indicated oligonucleotide primers. Coding sequences are represented by open bars; 5′ and 3′ noncoding sequences are represented by hatched bars. B, *Bgl* II; P, *Pst* I; S, *Sma* I; K, *Kpn* I; KB, kb. Arrows in lower section show direction and extent of sequences determined.

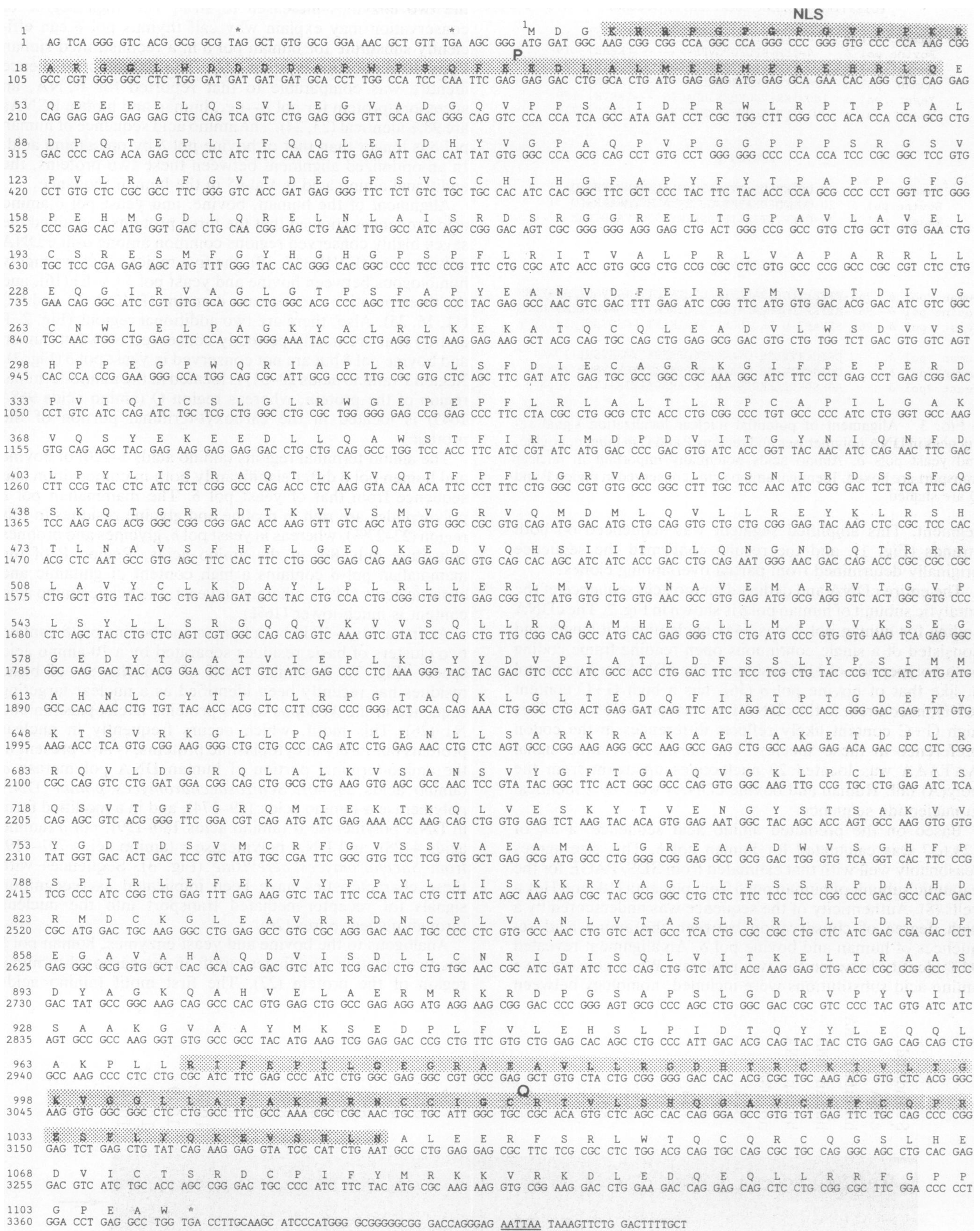

FIG. 2. Nucleotide and predicted amino acid sequence of human pol δ. In-frame stop codons are marked by *. The polyadenylylation signal is double underlined. The potential nuclear localization signal (NLS) (amino acids 4–19), region P (amino acids 20–51), and region Q (968–1044) are shaded. Conserved regions A–E (16) correspond to amino acids 126–153, 298–337, 465–561, 789–804, and 1047–1089, respectively; conserved regions I–VII (17, 18) correspond to amino acids 751–758, 581–621, 687–728, 368–409, 823–837, 654–671, and 806–811, respectively; and conserved exonuclease regions I, II, and III (11, 16, 25) correspond to amino acids 310–326, 393–409, and 504–519, respectively.

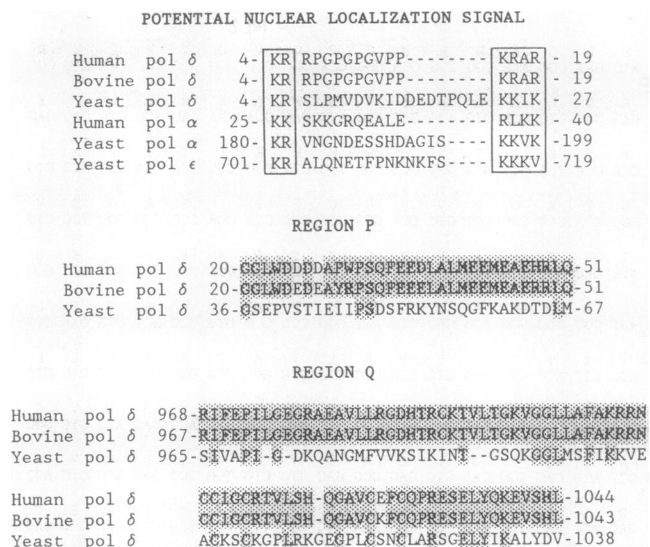

FIG. 3. Alignment of potential nuclear localization signal sequences in DNA polymerases and regions P and Q in human, bovine, and yeast pols δ. Amino acids potentially important in nuclear transport are boxed. Identical amino acids conserved in region P and Q are shaded.

segment. This amplified segment was sequenced on both strands (Fig. 1), and the results confirmed the sequence originally determined from partial overlapping clones.

The entire cDNA and deduced amino acid sequence of the catalytic subunit of human pol δ is shown in Fig. 2. The cDNA coding for human pol δ was 3443 nucleotides in length and consisted of a single continuous open reading frame coding for a polypeptide of 1107 residues. The cDNA for human pol δ, like that of bovine pol δ (16), has a high G+C content (65%), in contrast to the cDNA for yeast pol δ (38%) (7). This high G+C content likely reflects differences in the codon preference in these species. The polyadenylylation signal AATAAA was located 20 nucleotides upstream from the poly(A) tail. Human and bovine cDNAs were 89% identical in nucleotide sequence.

Based on the predicted amino acid sequence, a M_r of 123,657 was calculated for human pol δ. This size agrees reasonably well with that estimated from SDS/PAGE for the catalytic subunit of human pol δ from placenta (10) and HeLa cells (8). Authenticity of the sequence was underscored by a high degree of conservation between the amino acid sequences of human and bovine pol δ. An alignment revealed 94% identity between the two proteins. When conservative amino acid substitutions were included, homology between the two enzymes increased to 96%. This high degree of conservation may explain why calf thymus pol δ can efficiently substitute for human pol δ in a reconstituted simian virus 40 replication system (22). The extent of sequence identity was comparable to that reported for PCNA, an accessory protein for pol δ—i.e., human and mouse PCNAs are 98% identical (23, 24). The amino acid sequence of human pol δ is longer than that of bovine pol δ by one amino acid. In an optimized alignment between these two proteins, the additional amino acid is Val-65 in the human sequence.

Alignment of the human, bovine, and yeast pol δ amino acid sequences showed that the three proteins contained the seven highly conserved regions common among α-like DNA polymerases (I–VII) (17, 18), the five regions that are highly homologous between bovine and yeast pol δ (A–E) (16), and the three putative exonuclease domains (EXO I, II, and III) (11, 16, 25). Also, there are two additional regions (Fig. 2, P and Q) that are nearly perfectly conserved between human and bovine pol δ but are not conserved in yeast pol δ (Fig. 3). Region P (amino acids 20–51) is located in the amino-terminal region of the protein, whereas region Q (amino acids 968–1045) is located in the carboxyl-terminal portion of the protein.

The amino-terminal regions (amino acids 1–120) of bovine and human pol δ differ significantly both in composition and sequence from that of yeast pol δ. The mammalian pol δ polypeptides are rich in proline and glycine residues in this region (22–27%), whereas in yeast pol δ, glycines and prolines constitute only 8% of the amino acids. Region P of the mammalian pol δ contains a high content of glutamic and aspartic acids (34%), whereas in yeast, the acidic amino acid content is much lower (16%).

Amino acids 4–19 in both human and bovine pol δ contain two clusters of basic residues separated by a 10-amino acid spacer (Fig. 3). A similar motif of two clusters of basic residues has recently been identified as a nuclear targeting sequence in the *Xenopus laevis* proteins nucleoplasmin and N1 (26). This motif, which occurs frequently in nuclear proteins but rarely in cytoplasmic proteins, is also present in the amino-terminal portion of human DNA polymerase α (amino acids 25–40), *Schizosaccharomyces pombe* DNA polymerase α (amino acids 159–174), and in a modified form in DNA polymerase α (amino acids 180–199), pol δ (amino acids 4–28), and DNA polymerase ε (amino acids 701–719) from *Saccharomyces cerevisiae* (Fig. 3). Sequences with these characteristics have been implicated as recognition signals for receptor-mediated transport into the nucleus through nuclear pores (26).

Analogous to the bovine and yeast enzymes, human pol δ has two putative zinc-finger motifs in the carboxyl-terminal region of the protein (27). The first motif (amino acids

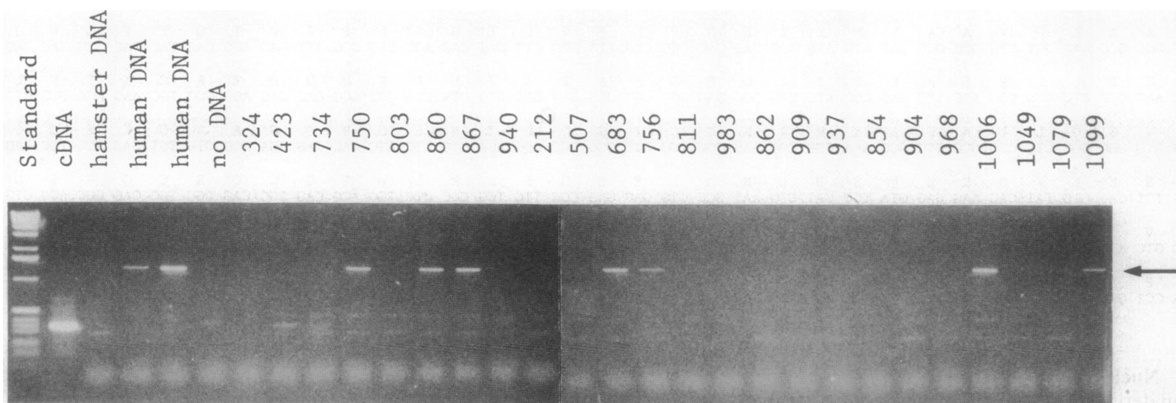

FIG. 4. Determination of chromosomal location by PCR analysis of DNA from human–hamster hybrid cell lines. Arrow identifies position of the characteristic 1.37-kb fragment from the human pol δ gene.

1012–1029) has the structure of Cys-Xaa_2-Cys-Xaa_{10}-Cys-Xaa_2-Cys, whereas the second motif (amino acids 1058–1076) has the structure of Cys-Xaa_2-Cys-Xaa_9-Cys-Xaa_4-Cys. The spacing between the two zinc fingers is perfectly conserved among the human, bovine, and yeast enzymes. However, the first zinc finger is present within region R, where low homology between the mammalian and yeast enzymes exists. In contrast, the second zinc finger is located within region E, a region noted to be highly conserved between bovine and yeast pol δ (16). Further studies are necessary to establish whether the putative zinc fingers are involved in protein–DNA or protein–protein interactions.

Chromosomal Location of the Gene for Human pol δ. Oligonucleotide primers HGF2 and HGR4 were used in a PCR reaction at an annealing temperature of 65°C to specifically amplify a segment of the human pol δ gene from total human genomic DNA. A major product of 1.37 kilobases (kb) was obtained under these conditions (Fig. 4). Cloning and partial DNA sequencing confirmed the authenticity of this fragment and further showed that the human pol δ gene contained at least two introns in this region, including a type I intron at Gly-1023 and a type II intron at Ser-1073. Therefore, this fragment was indicative of the presence of the pol δ gene. Under the same conditions, hamster DNA gave only faint bands 200 to 400 base pairs in length. PCR analysis of the DNA from a panel of 24 human–hamster hybrid cell lines showed that 7 cell lines produced the characteristic 1.37-kb fragment (cell lines 750, 860, 867, 683, 756, 1006, and 1099, Fig. 4), inferring that the gene for human pol δ must be present in a human chromosome common to these 7 cell lines but not in any of the chromosomes in the other 17 cell lines. Analysis of the chromosomal content of this panel showed complete concordance of the human pol δ gene with chromosome 19. These studies provide strong evidence for the conclusion that the human pol δ gene is located on chromosome 19. The gene coding for the catalytic polypeptide of human DNA polymerase α has been mapped to the X chromosome (28), whereas that for human PCNA has been localized to chromosome 20 (29). The mechanism by which these genes are coordinately expressed during the cell cycle, however, is not known.

We thank Dr. Joost Meijers for providing cellular RNA from HepG2 cells and Jeff Harris for technical assistance in oligonucleotide synthesis and DNA sequencing. We also thank Dr. Fred Hagen for helpful suggestions and discussion. This work was supported, in part, by Research Grants DK26206 and HL16919 from the National Institutes of Health.

1. Challberg, M. D. & Kelly, T. J. (1989) *Annu. Rev. Biochem.* **58**, 617–717.
2. Stillman, B. (1989) *Annu. Rev. Cell Biol.* **5**, 197–246.
3. Wang, T. S.-F. (1991) *Annu. Rev. Biochem.* **60**, 513–552.
4. So, A. G. & Downey, K. M. (1991) *Crit. Rev. Biochem. Mol. Biol.*, in press.
5. Lee, M. Y. W. T., Tan, C.-K., Downey, K. M. & So, A. G. (1984) *Biochemistry* **23**, 1906–1913.
6. Ng, L., Tan, C.-K., Downey, K. M. & Fisher, P. A. (1991) *J. Biol. Chem.* **266**, 11699–11704.
7. Boulet, A., Simon, M., Faye, G., Bauer, G. A. & Burgers, P. M. J. (1989) *EMBO J.* **8**, 1849–1854.
8. Syvaoja, J., Suomensaari, S., Nashida, C., Goldsmith, J. S., Chui, G. S. J., Jain, S. & Linn, S. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 6664–6668.
9. Goulian, M., Herrmann, S. M., Sackett, J. W. & Grimm, S. L. (1990) *J. Biol. Chem.* **265**, 16402–16411.
10. Lee, M. Y. W. T., Jiang, Y. Q., Zhang, S. J. & Toomey, N. L. (1991) *J. Biol. Chem.* **266**, 2423–2429.
11. Simon, M., Giot, L. & Faye, G. (1991) *EMBO J.* **10**, 2165–2170.
12. Tan, C.-K., Castillo, C., So, A. G. & Downey, K. M. (1986) *J. Biol. Chem.* **261**, 12310–12316.
13. Prelich, G., Tan, C.-K., Kostura, M., Mathews, M. B., So, A. G., Downey, K. M. & Stillman, B. (1987) *Nature (London)* **326**, 517–520.
14. Bauer, G. A. & Burgers, P. M. J. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 7506–7510.
15. Burgers, P. M. J. (1988) *Nucleic Acids Res.* **16**, 6297–6307.
16. Zhang, J., Chung, D. W., Tan, C.-K., Downey, K. M., Davie, E. W. & So, A. G. (1991) *Biochemistry*, in press.
17. Wong, S. W., Wahl, A. G., Yuan, P.-M., Arai, N., Pearson, B. E., Arai, K., Korn, D., Hunkapiller, M. W. & Wang, T. S.-F. (1988) *EMBO J.* **7**, 37–47.
18. Spicer, E. K., Rush, J., Fung, C., Reha-Krantz, L. J., Karam, J. D. & Konigsberg, W. H. (1988) *J. Biol. Chem.* **263**, 7478–7486.
19. Sambrook, J., Fritsch, E. G. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY), 2nd Ed.
20. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
21. Tait, J. F., Frankenberry, D. A., Maio, C. H., Killary, A. M., Adler, D. A. & Disteche, C. M. (1991) *Genomics* **10**, 441–448.
22. Tsurimoto, T., Melendy, T. & Stillman, B. (1990) *Nature (London)* **346**, 534–539.
23. Almendral, J. M., Huebsch, D., Blundell, P. A., Macdonald-Bravo, H. & Bravo, R. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 1575–1579.
24. Matsumoto, K., Moriuchi, T., Koji, T. & Nakane, P. K. (1987) *EMBO J.* **6**, 637–642.
25. Bernad, A., Blanco, L., Lazaro, J. M., Martin, G. & Salas, M. (1989) *Cell* **59**, 219–228.
26. Robbins, J., Dilworth, S. M., Laskey, R. A. & Dingwall, C. (1991) *Cell* **64**, 615–623.
27. Berg, J. M. (1990) *J. Biol. Chem.* **265**, 6513–6516.
28. Wang, T. S.-F., Pearson, B. E., Suomalainen, H. A., Mohandas, T., Shapiro, L. J., Schröder, J. & Korn, D. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 5270–5274.
29. Ku, D.-H., Travali, S., Calabretta, B., Huebner, K. & Baserga, R. (1989) *Somatic Cell Mol. Genet.* **15**, 297–307.