Proc. Natl. Acad. Sci. USA
Vol. 85, pp. 2489–2493, April 1988
Biochemistry

Human terminal deoxyribonucleotidyltransferase: Molecular cloning and structural analysis of the gene and 5′ flanking region

(human DNA polymerase/tissue-specific gene regulation)

LELA K. RILEY*, JENNIFER K. MORROW, MARY JO DANTON†, AND MARY SUE COLEMAN‡

Department of Biochemistry and the Lucille P. Markey Cancer Center, University of Kentucky Medical Center, Lexington, KY 40536-0084

Communicated by Mary Ellen Jones, December 21, 1987 (received for review October 21, 1987)

ABSTRACT **Human terminal deoxyribonucleotidyltransferase (nucleoside-triphosphate:DNA deoxynucleotidylexotransferase, EC 2.7.7.31) cDNA contains an open reading frame of 1530 base pairs (bp) corresponding to a protein containing 510 amino acids. The encoded protein is a template-independent DNA polymerase found only in a restricted population of normal and malignant prelymphocytes. To begin to investigate the genetic elements responsible for the tissue-specific expression of terminal deoxyribonucleotidyltransferase, genomic clones containing the entire human gene were isolated and characterized. Initially, cDNA clones were isolated from a library generated from the human lymphoblastoid cell line, MOLT-4R. A cDNA clone containing the entire coding region of the protein was used to isolate a series of overlapping clones from two human genomic libraries. The gene comprises 11 exons and 10 introns and spans 49.4 kilobases. The 5′ flanking region (709 bp) including exon 1 was sequenced. Several putative transcription initiation sites were mapped. Within 500 nucleotides of the translation start site, a series of promoter elements was detected. "TATA" and "CAAT" sequences, respectively, were found to start at nucleotides −185 and −204, −328 and −370, and −465 and −505. Start sites were found for a cyclic AMP-dependent promoter analog at nucleotide −121, an eight-base sequence corresponding to the IgG promoter enhancer (*cd*) at nucleotide −455, and an analog of the IgG promoter (*pd*) at nucleotide −159. These findings suggest that transcripts coding for terminal deoxyribonucleotidyltransferase may be variable in length and that transcription may be influenced by a variety of genetic elements.**

Expression of the template-independent DNA polymerase, terminal deoxyribonucleotidyltransferase (terminal transferase; nucleoside-triphosphate:DNA deoxynucleotidylexotransferase, EC 2.7.7.31) is tissue-specific. Activity is normally restricted to prelymphocytes of marrow and thymus (1, 2). Strikingly elevated levels of terminal transferase are detected in malignant prelymphocytes (3). The primary structure of terminal transferase is highly conserved. Comparisons of terminal transferase cDNAs isolated from calf, mouse, and human libraries reveal a sequence homology of >80% (4).

Terminal transferase can synthesize DNA without a template, and it has been postulated that this enzyme may play a role in creating somatic diversity at recombination junctions of variable, diversity, and joining regions of immunoglobulin and T-cell receptor genes (5). This hypothesis has been supported by data correlating the existence of *N* regions (small numbers of nucleotides not coded for by the chromosome) with the expression of terminal transferase mRNA (6) or protein (7) and with the insertion of the terminal transferase gene into a cell line that rearranges an immunoglobulin gene but lacks the enzyme (8). However, the precise mechanism by which this enzyme may function in this or other cellular processes is not understood.

To begin to approach some of these questions, we have elucidated the entire structure[§] of the human terminal transferase gene from a series of overlapping clones isolated from two genomic libraries constructed from germ-line DNA. In addition, we identified several putative regulatory elements in the 5′ flanking region of the gene.

MATERIALS AND METHODS

Materials. Radiolabeled nucleotides were obtained from New England Nuclear, and unlabeled nucleotides were from Pharmacia P-L Biochemicals. Enzymes were from New England Biolabs, International Biotechnologies (New Haven, CT), Collaborative Research (Waltham, MA), or Bethesda Research Laboratories. The human lymphoblastoid T-cell line MOLT-4R was a gift of J. Minowada (Okayama, Japan). Oligonucleotides were synthesized using an Applied Biosystems (Foster City, CA) DNA synthesizer.

MOLT-4R cDNA Library Construction. Total polysomal RNA and polyribosomes were isolated from 7 g of logarithmically growing MOLT-4R lymphoblasts (9, 10). Polysomal RNA was extracted (11) and poly(A)$^+$ RNA was isolated (12). Double-stranded cDNA was synthesized (13) by using modifications of Gubler and Hoffman (14) and was inserted into the *Pst* I site of pBR322 by homopolymer tailing; annealed recombinant DNA was used to transform *Escherichia coli* K802 (15).

Isolation and Analysis of Clones from cDNA and Genomic Libraries. The MOLT-4R cDNA library was used to transform *E. coli* DG75. Oligonucleotide labeling, prehybridization, and hybridization conditions were as described by Woods (16).

Two human genomic libraries were used. A fetal liver library was the gift of T. Maniatis (17), and a human leukocyte genomic library was from Clontech Laboratories (Palo Alto, CA). DNA probes were labeled with [α-^{32}P] dNTP (18). Membranes were prehybridized and hybridized at 42°C in a solution that included 50% formamide.

DNA from cDNA or genomic clones was characterized by restriction mapping, Southern blot analysis (19), and DNA

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.

Abbreviation: terminal transferase, terminal deoxyribonucleotidyltransferase.

*Present address: Department of Veterinary Pathology, College of Veterinary Medicine, University of Missouri, Columbia, MO 65211.

†Present address: Children's Hospital Research Foundation, University of Cincinnati, Cincinnati, OH 45229.

‡To whom reprint requests should be addressed.

§This sequence is being deposited in the EMBL/GenBank data base (Immunogenetics, Mountain View, CA, and Eur. Mol. Biol. Lab., Heidelberg) (accession no. J03243).

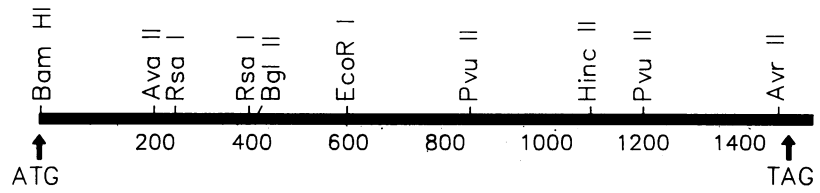

FIG. 1. Restriction map of human terminal transferase cDNA isolated from the MOLT-4R library.

sequencing of subclones in pBR322 (20) or in phage M13 vectors (21, 22).

Primer Extension and T4 DNA Polymerase Mapping of Transcription Start Sites. For reverse transcriptase reactions (23), two 30-mer primers corresponding to bases 18–47 and 43–72 in terminal transferase exon 1 were used. For T4 DNA polymerase analysis, the technique developed by Hu and Davidson (24) was used.

RESULTS

Isolation of Terminal Transferase cDNA Clones from the MOLT-4R Library. The MOLT-4R cDNA library was plated and screened (15). The ^{32}P-labeled DNA probe used was a synthetic 70-mer, constructed from the known cDNA sequence (25). From the original screening, 10 clones (5 full-length) were isolated. Clone identity was confirmed by restriction digests, Southern blot analysis, and DNA sequencing (Fig. 1).

Isolation and Mapping of the Human Terminal Transferase Gene. One of the largest cDNA clones [1770 base pairs (bp)] was radiolabeled and used to screen a human (fetal liver) genomic DNA library. From this primary screening, 14 overlapping clones were isolated. These were shown to span the region from the 3′ end of intron 1 through the 3′ end of exon 11.

Two additional library screenings were carried out, a repeat screening of the fetal liver library and a primary screening of a human leukocyte library. The probe used was a genomic DNA fragment that was the most 5′-terminal *Eco*RI restriction fragment. One clone from each library was found to span the entire 5′ end of the cDNA and to include nucleotide sequences of the regulatory region. The map was constructed by Southern blot analysis using terminal transferase cDNA fragments and synthetic oligonucleotides as probes (Fig. 2).

Gene Organization. The human terminal transferase gene spans 49.4 kilobases (kb) of DNA and contains 11 exons and 10 introns. Almost 90% of the exons were located in the 3′ half of the gene. The exon–intron junctions (Table 1) were deduced by comparison of genomic sequences with the terminal transferase cDNA sequence. The splice junctions are in good agreement with the consensus sequences (26).

The location and the sizes of the exon sequences are summarized in Table 1. The first intron comprises >50% of the gene length, spanning more than 29,000 bp.

5′ End of the Human Terminal Transferase Gene. We sequenced both complementary DNA strands 709 bp up-

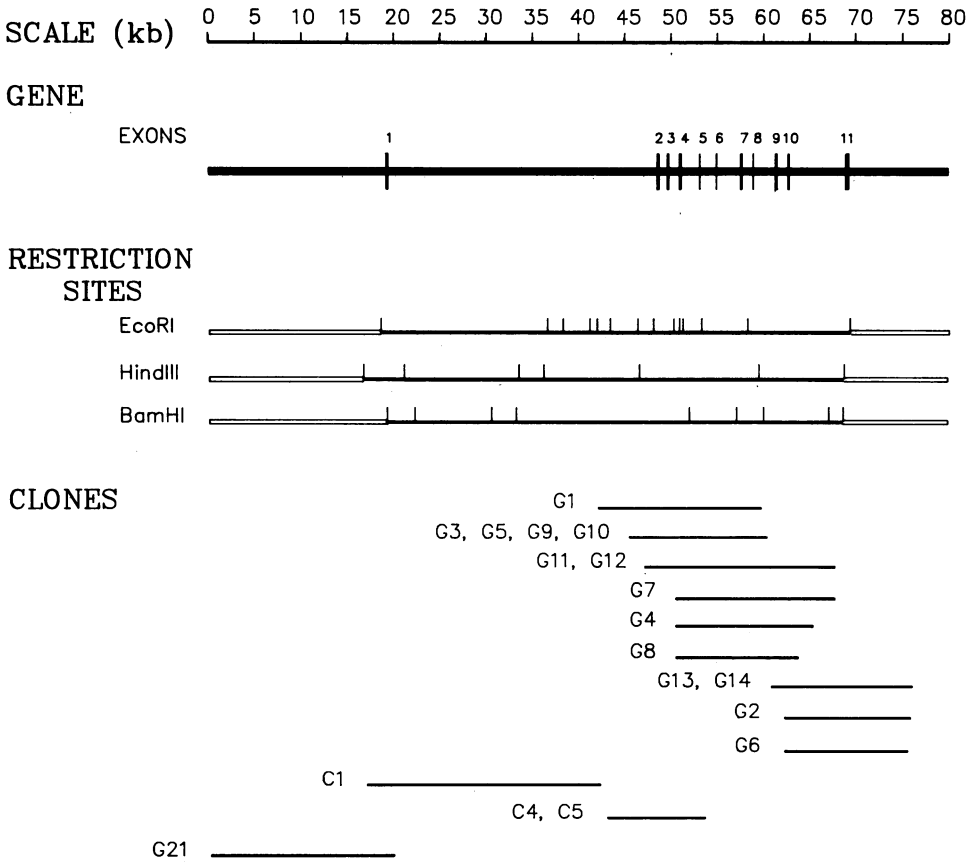

FIG. 2. Structure of the human terminal transferase gene. The open bars indicate flanking regions that have not been completely mapped. The locations and sizes of the 11 exons are indicated by vertical bars on the upper lines.

Table 1. Exon–intron splice-junction sequences in the human terminal transferase gene

Exon	Position in open reading frame	Size of exon, bp	Sequence of exon–intron junctions* 5′-splice donor....intron....3′-splice acceptor
1	1–203	203	GCTCAG gtaggacagc.....catttgcag TGATTC
2	204–378	175	CTTGTT gtagtgtcat.....aattgttaag GTGAGA
3	379–507	129	TTCACG gtaacggact.....tatcctgcag GATGCC
4	508–678	171	ATAGAG gtaagggtga.....tgtccattag GAGATT
5	679–750	72	TTCAAA gtaagtgatt.....tccattttag CTCTTT
6	751–874	124	AAGCAG gtaattgtct.....ggcctcttag GATTTC
7	875–1007	133	CCGGAG gtaattaact.....gaaaattgag GGGTAA
8	1008–1113	106	AAGAAG gtgagaagaa.....ttcttttcag GGATTA
9	1114–1362	249	CCTCGG ttgccgccgg.....tcctatccag CAGTTT
10	1363–1446	85	ACCAAG gtacagttcc.....tgttttcgag AGGATA
11	1447–1746	300†	GCCTAG

*Upper-case letters signify exon sequence, and lower-case letters signify intron sequence.
†This exon contains 84 bp of the 3′ end of the open reading frame.

stream from the translation ATG start codon (Fig. 3). A number of putative regulatory elements were identified (Table 2). Three "TATA" boxes (30) coupled with "CAAT" sequences (30) were found to start at nucleotides −465, −328 and −185. A region with a high degree of homology with the core sequence found in cyclic AMP promoters (31) was identified as starting at nucleotide −121. In addition, we have identified two sequences that are implicated in promoters in the immunoglobulin genes: a *cd* octamer element starting at nucleotide −455 and an analog of the *pd* element starting at nucleotide −159 (32).

The DNA sequence at the 5′ terminus of the terminal transferase gene was analyzed for dyad symmetry by using parameters set by Tinoco *et al.* (27), and three overlapping sequences between −108 and −158 were identified. Potential secondary structures could be generated between residues −158 and −108 (ΔG = −15.3 kcal/mol), −149 and −122 (ΔG = −17.7 kcal/mol), and −136 and −110 (ΔG −11.4 kcal/mol). The significance of such putative structures is not clearly understood.

The results of primer extension analyses are shown in Fig. 4. There was microheterogeneity in the DNA products. Three sites (nucleotides −100, −200, and −400) appear to be plausible, since the TATA and CAAT sequences are within about 100 nucleotides of these regions (28, 29).

The T4 DNA polymerase "primer extension" also was included because it has been designed for the structural elements we found in the gene: a long distance between exons 1 and 2 as well as a potentially long 5′ untranslated region. These experiments revealed an untranslated DNA

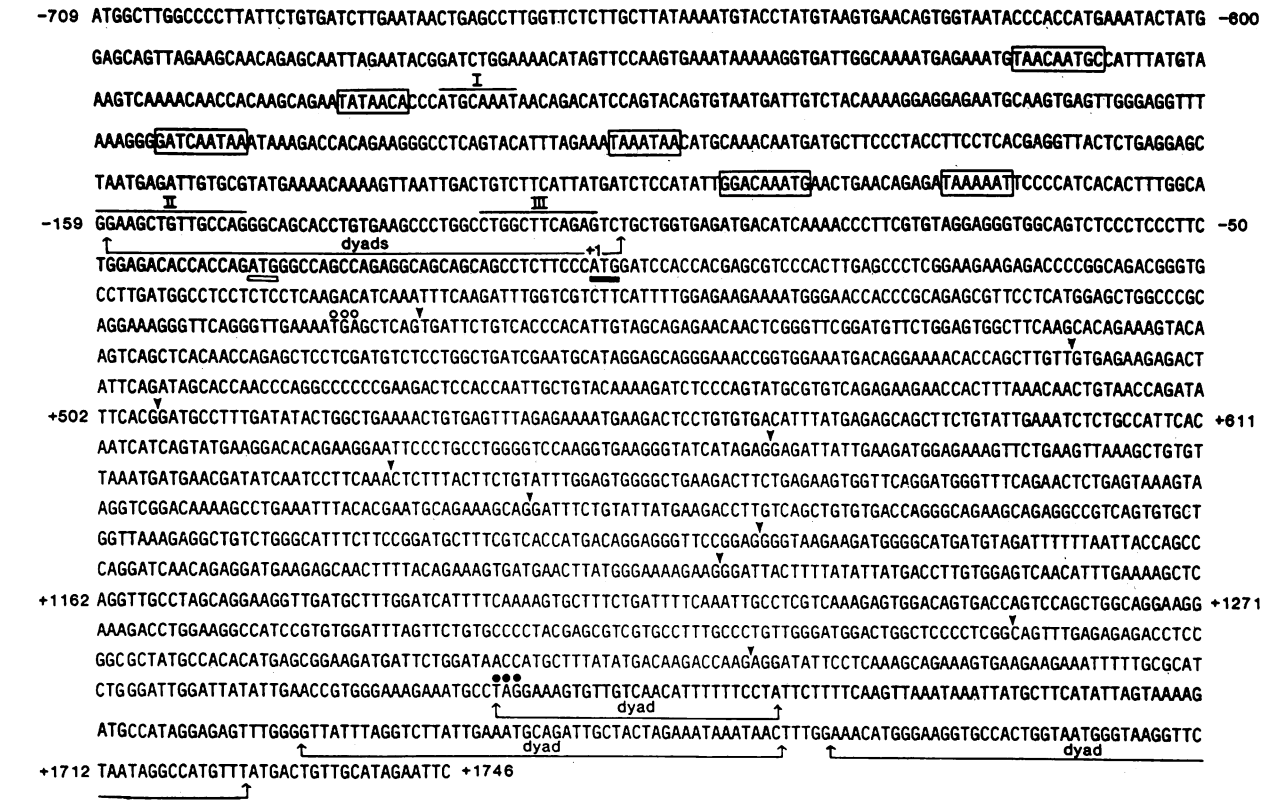

FIG. 3. The sequence of the 5′ regulatory region and the exons of the human terminal transferase gene. The ATG translation start codon is underlined (thick bar) and designated +1, and its termination is noted with ●●●. The TATA and CAAT sequences are boxed. The potential regulatory sequences (see Table 2) are designated by I (*cd*), II (*pd*), and III (cAMP promoter). The out-of-frame ATG is underlined by an open bar, and its termination codon is designated by ○○○. Areas of dyad symmetry are underlined and enclosed with arrows. The exon splice sites are marked by ▼.

Table 2. Location and identification of potential regulatory elements

Regulatory element	5′ Sequence 3′*	Location†
TATA	TATA$^{A}_{T}$A$^{A}_{T}$	Consensus (27)
TATA	TATA**C**A	−465 to −459
TATA	TA**A**ATAA	−328 to −322
TATA	TA**A**AAAT	−185 to −179
CAAT	GG$^{C}_{T}$CAATCT	Consensus (27)
CAAT	T**A**ACAAT**GC**	−508 to −500
CAAT	G**A**TCAAT**AA**	−373 to −365
CAAT	GG**A**CAA**ATG**	−207 to −199
Cyclic AMP promoter	CTTACGTCAGAG	Core sequence (28)
Cyclic AMP promoter	CT**GG**C**T**TCAGAG	−121 to −110
IgG promoter enhancer (*cd*)‡	ATGCAAAT	Core sequence (29)
IgG promoter enhancer (*cd*)	ATGCAAAT	−455 to −448
IgG promoter enhancer (*pd*)§	TGCA$^{G}_{C}$CTGTGNCCAG	Core sequence (29)
IgG promoter enhancer (*pd*)	G**G**A**A**GCTGT**T**GCCAG	−159 to −145

*The underlined bases in the terminal transferase gene sequence are mismatched from the consensus sequence.
†The adenosine of the ATG translation start site is designated +1. Numbers in parentheses are refs.
‡This octamer element has been shown to direct lymphocyte-specific RNA synthesis when it is within 70 bp of the start site of transcription.
§This 15-bp element has been associated with promoters in immunoglobulin genes.

stretch of 334–378 bp and placed potential transcription start sites at around nucleotide −400. We also detected start sites that would generate shorter untranslated regions, consistent with the primer extension analyses described above.

The genomic sequences were compared with two published sequences of human terminal transferase cDNAs. We set the reading frame by comparing the protein sequence we have determined with the human nucleotide sequence. Genomic sequences from nucleotide −89 through 1746 were identical to the human cDNA sequence published by Chang and Bollum (33) except at nucleotide 1358, where our sequence has an extra triplet (CTC) immediately preceding the final two bases of the 3′ end of exon 9, and at nucleotide 1363, where we have a CAG at the 5′ end of exon 10 (see Fig. 3 and Table 1). Comparison of our DNA sequence with that

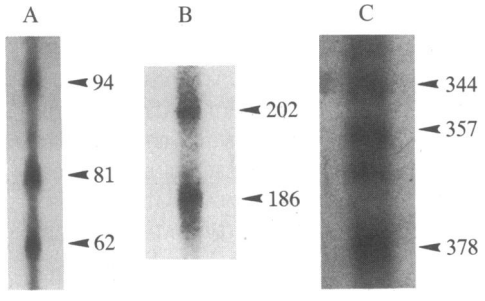

FIG. 4. Primer extension (*A* and *B*) and T4 DNA polymerase mapping (*C*) of transcription initiation sites. Data from three representative experiments are shown to illustrate multiple transcription start sites. The primers used for these experiments were: an oligonucleotide corresponding to bases 43–72 of exon 1 (*A*), an oligonucleotide corresponding to bases 18–47 of exon 1 (*B*), and phage M13 universal primer (*C*).

of Koiwai *et al.* (34) reveals several differences. The open reading frame of our DNA is 1530 bp. In contrast, they report an open reading frame of 1557 bp. The difference in the positioning of the functional ATG is due to an ATG at positions −34 to −32 in the sequence shown in Fig. 3. However, this ATG is out-of-frame in our cDNA and the genomic sequence and results in translation of a short nonhomologous peptide.

The sequence context of the ATG translation start site for terminal transferase is CCTCTTCCC**ATG**G. On the 3′ side of the ATG codon, the flanking nucleotide is a guanosine, which is preferred in vertebrate mRNAs (35).

3′ End of the Human Terminal Transferase Gene. Exon 11 contains the translation termination codon for terminal transferase at base 1531. Three regions of dyad symmetry were found: between positions 1531 and 1559 (ΔG = −5 kcal/mol), between 1622 and 1669 (ΔG = −4.7 kcal/mol), and between 1675 and 1726 (ΔG = −8.5 kcal/mol). These regions are clustered at the end of the open reading frame and may be used as part of termination events.

DISCUSSION

Terminal transferase is one of five major DNA polymerases found in vertebrate cells. Like DNA polymerases α, β, and γ and the recently described δ polymerase (36), terminal transferase catalyzes the polymerization of deoxyribonucleotides. However, terminal transferase does not copy a template and is restricted to subpopulations of lymphocytes and leukemic cells. Whether all DNA polymerases evolved from a common primordial polymerase is not yet known. Preliminary comparative studies of terminal transferase and DNA polymerase β protein sequences have revealed interesting features (37). A stretch of 174 amino acids in both proteins has been shown to contain 124 residues with identical favored and conservative amino acid substitutions. This similarity may reflect a common precursor gene or may reflect a protein structural requirement for the common nucleotide and DNA binding functions.

The human terminal transferase gene structure presented in this paper exhibits several unusual features. The total size of this gene (in relation to the 2.0-kb cDNA) is very large. Exons constitute approximately 5% of the gene, and their distribution is highly asymmetric. The first intron comprises more than half of the total gene size.

The question of whether terminal transferase mRNAs have variable-length leaders or uniformly short or long leaders is not resolved. Two primer-extension techniques yielded a pattern of putative sites. Three-size classes of 5′ untranslated sequences corresponding to 100, 200, and 400 nucleotides were repeatedly detected with poly(A)$^+$ RNA preparations isolated from several cell lines and leukemic cell samples. At each putative leader sequence, we detected a TATA box coupled with a CAAT sequence analog. One of these is preceded by an octamer sequence that has been found in the −100 region of all heavy chain variable region genes (32) and which has recently been demonstrated to be sufficient for lymphoid-specific gene expression (38). The octamer element has been shown in model systems to function in fairly close proximity to the transcriptional start site and to resemble CCAAT and G+C-rich sites found as part of other promoters. The uniqueness of the octamer element is that it shows tissue specificity. An additional promoter element, homologous to those found in genes whose expression is regulated by cAMP (31), was detected within 120 nucleotides of translation initiation.

These putative regulatory elements suggested that the untranslated region of the terminal transferase mRNA may be quite long. Peterson *et al.* (25) have reported the isolation of a terminal transferase cDNA clone that contains 328 bp upstream from the first amino acid codon of the protein.

However, comparison of our 5'-gene sequence with that cDNA sequence (33) revealed that after position −89, the sequences diverged completely. There is no evidence of a splice junction in the genomic sequence at this site to account for the difference. However, a stretch of 8 cytidine and 14 guanosine residues, which could have been derived from the homopolymer tailing, did occur in the published cDNA at the position of sequence divergence.

RNA blot analyses in our laboratory yield a major species of terminal transferase mRNA of about 2100 nucleotides. Calculations based on the length of the 3' untranslated region and the protein coding sequence give a 5' leader of about 250 nucleotides. Even given these observations, it is important to note that extremely long 5' mRNA leader sequences are very unusual (39). In most vertebrate mRNAs, the distance between transcription and translation start signals is usually between 20 and 80 nucleotides. Further, within the 400 nucleotides 5' to the terminal transferase translation start site, two out-of-frame ATG triplets are observed, in contrast to most mRNAs. Additional experimentation using nuclease S1 mapping and a variety of hybridization conditions will be necessary to resolve this question.

Inspection of the nucleotide sequence upstream from the putative translation initiation AUG (A is designated as +1) codon reveals several interesting features. The optimal sequence for initiation by 40S ribosomes has been found to be ANNAUGG. The most highly conserved position in this motif is a purine at −3. Of 699 mRNA sequences surveyed, 97% have a purine at that position (35). In addition, guanosine is preferred in positions −6 and −9. In the +4 position, guanosine is the preferred nucleotide. Of these general rules, terminal transferase exhibits the preferred nucleotide (guanosine) only in the +4 position; at −3, −6 and −9 positions, cytidine residues are present. The sequence found in the terminal transferase mRNA is shared by <2% of vertebrate mRNAs (35). Site-directed mutagenesis experiments have confirmed the contributions of the −3- and −6-position purines (40) in several mRNAs. Based on these data, the terminal transferase AUG would be designated as relatively "weak" in translation efficiency.

An additional rare feature of the terminal transferase mRNA sequence is the occurrence of the upstream nonfunctional ATG codon starting at position −34. The presence of upstream ATG codons has been demonstrated experimentally to be inhibitory to translation (40). This ATG also occurs in an unfavorable sequence context and is followed by a terminator codon starting at nucleotide +195 in the coding sequence. Whether translation of this "minicistron" (as a protein 76 amino acids in length) occurs in addition to translation of the major open reading frame for terminal transferase mRNA is not known.

The gene structure of a human non-template-dependent DNA polymerase that exhibits tissue-specific expression has now been elucidated. Not surprisingly, the regulatory region is quite different from the sequences found upstream from the gene for DNA polymerase β, a constituitively expressed enzyme that exhibits no tissue specificity (41). As the complete structure of the gene for DNA polymerase β and for other vertebrate DNA polymerases becomes available, it will be possible to search for common genetic features that may reveal the evolutionary origins of these proteins.

We are grateful for the technical assistance of Mr. John May and for the secretarial assistance of Mrs. Sara Thompson. This work was supported by Grant CA19492 from the National Cancer Institute and by Predoctoral Fellowship T32CA09509 from the National Institutes of Health to M.J.D.

1. Chang, L. M. S. (1971) *Biochem. Biophys. Res. Commun.* **44,** 124–131.
2. Coleman, M. S., Hutton, J. J., deSimone, P. & Bollum, F. J. (1974) *Proc. Natl. Acad. Sci. USA* **71,** 4404–4408.
3. Coleman, M. S. & Hutton, J. J. (1981) in *The Leukemic Cell: Methods in Hematology,* ed. Catovsky, C. (Churchill Livingstone Press, London), Vol. 2, pp. 203D–219D.
4. Koiwai, O., Yokota, T., Kageyama, T., Hirose, T., Yoshida, S. & Arai, K. (1986) *Nucleic Acids Res.* **14,** 5777–5792.
5. Blackwell, T. K. & Alt, F. W. (1984) *Cell* **37,** 105–112.
6. Yancopoulos, G. D., Blackwell, T. K., Suh, H., Hood, L. & Alt. F. W. (1986) *Cell* **44,** 251–259.
7. DeSiderio, S. V., Yancopoulos, G. D., Paskind, M., Thomas, E., Boss, M. A., Landau, N., Alt, F. W. & Baltimore, D. (1984) *Nature (London)* **311,** 752–755.
8. Landau, N. R., Schatz, D. G., Rosa, M. & Baltimore, D. (1987) *Mol. Cell. Biol.* **7,** 3237–3243.
9. Berger, S. L. & Birkenmeier, C. S. (1979) *Biochemistry* **18,** 5143–5149.
10. Kraus, J. P. & Rosenberg, L. E. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 4015–4019.
11. Brown, A. J. P., Leibold, E. A. & Munro, H. N. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 1265–1269.
12. Aviv, H. & Leder, P. (1972) *Proc. Natl. Acad. Sci. USA* **69,** 1408–1412.
13. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
14. Gubler, U. & Hoffman, B. J. (1983) *Gene* **25,** 263–269.
15. Hanahan, D. & Meselson, M. (1980) *Gene* **10,** 63–67.
16. Woods, D. (1984) *Focus* **6,** 1–3.
17. Lawn, R. M., Fritsch, E. F., Parker, R. C., Blake, G. & Maniatis, T. (1978) *Cell* **15,** 1157–1174.
18. Feinberg, A. P. & Vogelstein, B. (1983) *Anal. Biochem.* **132,** 6–13.
19. Southern, E. M. (1975) *J. Mol. Biol.* **98,** 503–517.
20. Chen, E. Y. & Seeburg, P. H (1985) *DNA* **4,** 165–170.
21. Sanger, F., Nicklen, S. & Coulsen, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
22. Yanisch-Perron, C., Vieira, J. & Messing, J. (1985) *Gene* **33,** 103–119.
23. Jahnsen, T., Hedin, L., Kidd, V. J., Beattie, W. G., Lohmann, S. M., Walter, U., Durica, J., Schulz, T. Z., Schlitz, E., Browner, M., Lawrence, C. B., Goldman, D., Ratoosh, S. L. & Richards, J. S. (1986) *J. Biol. Chem.* **261,** 12352–12361.
24. Hu, M. C.-T. & Davidson, N. (1986) *Gene* **42,** 21–29.
25. Peterson, R. C., Cheung, L. C., Mattaliano, R. J., White, S. T., Chang, L. M. S. & Bollum, F. J. (1985) *J. Biol. Chem.* **260,** 10495–10502.
26. Cech, T. R. (1983) *Cell* **34,** 713–716.
27. Tinoco, I., Borer, P. N., Dengler, B., Levine, M. D., Uhlenbeck, O. C., Crothers, D. M. & Gralla, J. (1973) *Nature (London) New Biol.* **246,** 40–41.
28. Corden, J., Wasylyk, B., Buchwalder, A., Sassone-Corsi, P., Kedinger, C. & Chambon, P. (1980) *Science* **209,** 1406–141.
29. Benoist, C., O'Hare, K., Breathnach, R. & Chambon, P. (1980) *Nucleic Acids Res.* **8,** 127–142.
30. Breathnach, R. & Chambon, P. (1981) *Annu. Rev. Biochem.* **50,** 349–383.
31. Short, J. M., Wynshaw-Boris, A., Short, H. P. & Hanson, R. W. (1986) *J. Biol. Chem.* **261,** 9721–9726.
32. Falkner, F. G. & Zachau, H. G. (1984) *Nature (London)* **310,** 71–74.
33. Chang, L. M. S. & Bollum, F. J. (1986) *Crit. Rev. Biochem.* **21,** 27–52.
34. Koiwai, O., Kaneda, T. & Morishita, R. (1987) *Biochem. Biophys. Res. Commun.* **144,** 185–190.
35. Kozak, M. (1987) *Nucleic Acids Res.* **15,** 8125–8148.
36. Crute, J. J., Wahl, A. F. & Bambara, R. A. (1986) *Biochemistry* **25,** 26–36.
37. Matsukage, A., Nishikawa, K., Ooi, T., Seto, Y. & Yamaguchi, M. (1987) *J. Biol. Chem.* **262,** 8960–8962.
38. Wirth, T., Staudt, L. & Baltimore, D. (1987) *Nature (London)* **329,** 174–178.
39. Sazar, S. & Schimke, R. T. (1986) *J. Biol.Chem.* **261,** 4685–4690.
40. Kozak, M. (1986) *Cell* **44,** 283–292.
41. Yamaguchi, M., Hirose, F., Hayashi, Y., Nishimoto, Y. & Matsukage, A. (1987) *Mol. Cell. Biol.* **7,** 2012–2018.