The Human Prothymosin α Gene Family Contains Several Processed Pseudogenes Lacking Deleterious Lesions

RICHARD E. MANROW, ALVARO LEONE,[1] MARC S. KRUG,[2] WILLIAM H. ESCHENFELDT,[3] AND SHELBY L. BERGER

Section on Genes and Gene Products, Laboratory of Biochemistry, National Cancer Institute, Bethesda, Maryland 20892

Received September 12, 1991; revised February 20, 1992

The six members of the human prothymosin α gene family have been cloned and sequenced. One gene (PTMA) contains introns and appears to be the source of all isolated prothymosin α cDNAs. The remaining five genes are processed pseudogenes. Four of them have consensus TATA elements upstream of sequences nearly identical to the transcriptional start region of the intron-containing gene. Those four genes also contain open reading frames coding for proteins closely related to prothymosin α. In two of the pseudogenes, PTMAP2 and 5, the encoded proteins differ from the product of the parental gene at only two and four locations, respectively. The fifth pseudogene (PTMAP1) encodes a different protein owing to an upstream translational initiation start site and multiple deletions and insertions. Because the potential for expression exists in this system, a search for pseudogenomic transcripts was undertaken using the polymerase chain reaction to amplify reverse transcripts of mRNAs from many human tissues and bulk DNA from several human cDNA libraries. Evidence for pseudogenomic transcripts was not obtained. Therefore, we conclude that the human prothymosin α gene family contains only one functional gene. © 1992 Academic Press, Inc.

INTRODUCTION

Prothymosin α is a highly acidic protein that was once believed to be a precursor for putative thymic hormones (Low and Goldstein, 1985) and, later, a secreted thymic hormone itself (Haritos *et al.*, 1985). However, subsequent work revealed that prothymosin α gene expression occurs ubiquitously in mammalian cells (Eschenfeldt and Berger, 1986; Gomez-Marquez *et al.*, 1989) and that the product, apparently without proteolytic processing, is translocated to the nucleus (Manrow *et al.*, 1991). Recently, we have demonstrated a direct correlation between elevated levels of prothymosin α gene activity and increased levels of cellular proliferation (Eschenfeldt and Berger, 1986). Furthermore, we have shown that the division of human myeloma cells is inhibited by treatment with antisense deoxyribonucleotide oligomers complementary to prothymosin α mRNA (Sburlati *et al.*, 1991). As a consequence, we propose a vital role for prothymosin α in the processes leading to cell division (Eschenfeldt and Berger, 1986; Sburlati *et al.*, 1991). This view is supported by experiments in which activation of c-*myc* in quiescent cells, in the absence of protein synthesis, stimulated transcription of prothymosin α genes (Eilers *et al.*, 1991).

As part of our continuing study of prothymosin α function, we have isolated the members of the human prothymosin α gene family. Early studies indicated that the human genome contained more than one prothymosin α-like sequence capable of hybridizing with a full-length cDNA (Eschenfeldt and Berger, 1986). Ultimately, six genes were identified and cloned[4] (Eschenfeldt *et al.*, 1989). Characterization of these genes suggested that only one contained introns, raising the possibility that the remaining genes were processed pseudogenes (Eschenfeldt *et al.*, 1989). This interpretation was supported by extensive sequencing of the intron-containing gene and partial sequencing of the 5′ ends of the putative pseudogenes. Clearly, further information was required to resolve the issue.

We report here that four of the five intronless genes are indeed processed pseudogenes. Although the extreme 3′ end of the fifth gene has not been cloned and sequenced, we present evidence that it, too, is likely to be a processed gene. The sequences detailed here, which include the entire prothymosin α coding region for each of the genes, indicate that four of the five pseudogenes possess relatively intact reading frames and encode prothymosin α-like polypeptides. Furthermore, at least one TATA-like sequence element (Breathnach and Chambon, 1981; Lewin, 1985) has been located at the 5′ end of each gene, a Kozak consensus sequence (Kozak, 1986, 1987) has been found immediately upstream of each potential translational initiation codon, and a conserved polyadenylation signal has been identified in all cases for which sequence information exists. The tanta-

[1] Present address: Laboratory of Pathology, National Cancer Institute.

[2] Present address: Department of Medicine, Division of Allergy and Immunology, University of Tennessee Memphis, Memphis TN 38163.

[3] Present address: Amoco Technology, P.O. Box 400, Naperville, IL 60566.

[4] The nucleotide sequences reported in this paper have been submitted to the GenBank/EMBL Data Bank under Accession Nos. J04797–J04802.

0888-7543/92 $5.00

lizing possibility that one or more of these pseudogenes may be expressed has been investigated, but convincing evidence has not been obtained. However, given that the intron-containing gene is alternatively spliced and that one of the transcripts represents a maximum of only 0.5% of the total in several cDNA libraries, the negative evidence may simply reflect our limits of detection (Manrow and Berger, unpublished).

MATERIALS AND METHODS

Prothymosin α genes. The isolation of the human prothymosin α gene family members[5] has been described previously (Eschenfeldt *et al.*, 1989). In brief, the functional gene (PTMA) and pseudogenes 1–4 (PTMAP1–4) were isolated from two human cosmid libraries (~2 × 10^6 clones). Pseudogene 5 (PTMAP5), the one remaining member, was not found among the cosmids, but was selected, instead, from a library of 2- to 3-kb *Eco*RI fragments of human DNA inserted into pGEM3 (Promega, Madison, WI) and cloned in *Escherichia coli* HB101.

DNA sequencing. Double-stranded DNA was sequenced (Sanger, 1981) using either the GemSeq K/RT system (Promega) or the Sequenase Version 2.0 system (U.S. Biochemicals, Cleveland, OH) in the presence of deoxyadenosine 5′-[α-^{35}S]thiotriphosphate (Amersham, Arlington, IL). Regardless of which system was used, the template DNA was denatured with NaOH as outlined in the procedures accompanying the GemSeq K/RT kit. Both the Klenow fragment of *E. coli* polymerase I (Promega) and avian myeloblastosis virus reverse transcriptase (Molecular Genetic Resources, Tampa, FL) were employed in GemSeq K/RT sequencing reactions. Initially, cosmid DNAs were sequenced directly using custom oligomers (OCS, Denton, TX) complementary to prothymosin α cDNA sequences; later, genes or fragments of genes were subcloned into the vectors pGEM3, pGEM3Z, pGEM3Zf(+), or pGEM3Zf(−) (Promega) and sequenced using the T7 or SP6 primers included in the GemSeq K/RT kit, or custom primers (OCS or Synthecell, Rockville, MD). Each annealing mixture contained 1–2 μg of plasmid DNA and 30 ng of primer in a final volume of 10 μl. In some cases, single-stranded DNA obtained with the use of the pGEM Single Strand system (Promega) was sequenced using Sequenase. Reaction products were analyzed in 8 or 10% polyacrylamide-urea gels, which were subsequently fixed and dried according to the GemSeq K/RT kit instructions. Dried gels were exposed to XAR-5 X-ray film at −80°C. Ambiguities in the resulting sequence ladders were resolved by sequencing the opposite strand, rerunning the sequencing reactions with a different enzyme, or analyzing the reaction products in a different gel. In a few cases, particular sequences were resequenced on one or both strands with the Multi-Pol DNA Sequencing System (Clontech, Palo Alto, CA). Autoradiographs of sequencing gels were read by at least two individuals working independently.

Data management. DNA sequences were recorded from autoradiographs of gels both manually and using a Gel Reader (International Biotechnologies, New Haven, CT) interfaced with an IBM PC XT computer. PC Gene (Intelligenetics, Mountain View, CA) programs were used as follows: READGEL to enter data; SEQIN to compile subsequences; RESTRI to identify restriction enzyme sites for subcloning of fragments; and TRANSL to recognize open reading frames in the compiled sequences. Comparison of nucleic acid sequences was performed using NUCALN (IBM PC Version 2.0), developed by D. J. Lipman and W. J. Wilbur (National Institutes of Health, Bethesda, MD). Final sequence alignments were constructed using a Macintosh IIcx computer and WORDPERFECT (Wordperfect Corporation, Orem, UT).

Analysis of mRNAs and cDNA libraries by means of the polymerase chain reaction. Bulk RNA containing mRNA from human placenta, liver, thyroid, striated muscle, colon, ovarian carcinoma, stomach, and a kidney cell line were kind gifts from Drs. Mark Sobel and William Linehan of the National Cancer Institute. They were reverse transcribed using a 1st Strand Synthesis Kit supplied by Stratagene (San Diego, CA). Our procedure, derived from instructions included in the kit, was as follows: 1 μg of total RNA from each tissue was annealed with 300 ng of oligo(dT) primer and 135 ng of PsG down primer (defined below) in 35 μl of water at 65°C for 5 min. As a control, the same components from each tissue were supplemented with 20 μg/ml of pancreatic ribonuclease A during the annealing step. The 16 samples were then cooled slowly to room temperature and reverse transcribed for 1 h at 37°C by the addition of 15 μl containing 20 units of Moloney murine leukemia virus reverse transcriptase and dithiothreitol, RNase Block II, deoxyribonucleoside triphosphates, and concentrated buffer as described in the kit.

Clonetech was the source of two human cDNA libraries prepared in λgt11 with mRNA isolated from normal skin fibroblasts (primary cultures) and IM9 transformed human myeloma cells. A human teratocarcinoma cDNA library cloned into λgt10 was a gift from Dr. Maxine Singer of the Carnegie Institute, Washington, D.C. (Skowronski *et al.*, 1988). DNA from each library was isolated and purified using a Qiagen λ kit (Pack 500) and instructions from the manufacturer (Studio City, CA).

The polymerase chain reaction (PCR) was carried out with either 5 μl of crude first-strand cDNA or 0.5–1 μg of library DNA and two oligomers; the upstream primer, PsG up, was identical to residues 152–172 of the gene, whereas the downstream primer, PsG down, was complementary to residues 486–507 (see Fig. 1). Both oligomers also included a 5′ extension bearing a spacer triplet (CCA) followed in the 3′ direction by a *Kpn*I site for subsequent cloning. Amplification was carried out in a volume of 100 μl using a Perkin-Elmer/Cetus DNA Thermocycler (Norwalk, CT) and buffers and instructions supplied by the manufacturer. The first cycle was performed at 94°C for 30 s, 55°C for 15 s, and 72° for 1 min, whereas in subsequent cycles (either 24 or 49) the incubation at 94°C was reduced to 15 s. To eliminate contaminating DNA, water was boiled for at least 30 min and passed through 0.45-μm Nalgene disposable filters, and sterile reaction tubes were vigorously shaken with 1 *M* HCl for 30 min, drained, and allowed to evaporate to dryness in a fume hood. Dedicated pipets and barrier tips were also employed.

The absence of DNA contamination in the RNAs and in the 1st Strand kit components was established by amplifying the reverse transcription products of the ribonuclease-treated samples described above. Since pancreatic ribonuclease A itself could also have been a source of DNA, it too was tested. Samples containing oligomers but no template were also subjected to PCR.

The library DNAs posed additional problems because PCR products obtained from cloned pseudogenomic cDNA could not readily be distinguished from the products of inadvertently cloned genomic DNA. A search for unwanted genomic DNA in the libraries was undertaken using the PCR reaction with oligomers specific for introns 2 and 3 of the functional prothymosin α gene. A positive result would indicate the presence of prothymosin α genomic DNA in the library and, by inference, the presence of other genomic DNAs. Library DNA was also amplified using an oligomer complementary to residues 312–332 of the gene (Fig. 1) together with an upstream oligomer identical to the 5′ flank of either PTMAP4 or 5 and an oligomer identical to residues 1042–1062 of the gene together with an oligomer directed toward the 3′ flank of either PTMAP2 or 3. The latter experiment was designed to expose contamination of the cDNA libraries with genomic DNA flanking the prothymosin α pseudogenes: if transcription of a pseudogene had been initiated at or very near position 1 and terminated near position 1230, such a PCR experiment would fail. A positive result with a particular oligomer pair would implicate a specific pseudogenomic DNA as a contaminant in the library. All oligomers were tested

[5] In accordance with the requirements of the Human Genome Mapping Workshop, the members of the prothymosin α gene family have been renamed as follows: gene 28, prothymosin, alpha (PTMA); gene sequence 26, prothymosin, alpha pseudogene 1 (PTMAP1); gene sequence 32, prothymosin, alpha pseudogene 2 (PTMAP2); gene sequence 34, prothymosin, alpha pseudogene 3 (PTMAP3); gene sequence 112, prothymosin, alpha pseudogene 4 (PTMAP4); and gene sequence 150, prothymosin, alpha pseudogene 5 (PTMAP5).

for their ability to serve as primers for PCR by amplifying ~10 ng of the relevant cloned pseudogenomic DNA templates or by amplifying ~10 ng of the cDNA clone derived from the functional gene; the resultant fragments were the correct sizes, 260–510 bp depending on which DNA and oligomers were tested.

Amplified fragments were analyzed electrophoretically in 2.5% agarose gels and, after cleavage with restriction enzymes, in either 2.5% agarose or 20% polyacrylamide gels with Tris–borate–EDTA buffer in both cases. Fragments were retrieved from gels by means of a modification of the freeze–squeeze method, with the squeezing step performed in a centrifuge (Ogden and Adams, 1987). The DNA was recovered by ethanol precipitation.

RESULTS AND DISCUSSION

Human Prothymosin α Genes

The six members of the human prothymosin α gene family have been isolated (Eschenfeldt *et al.*, 1989). Restriction enzyme analysis and Southern blotting (Southern, 1975) using probes derived from the 5′ and 3′ ends of a full-length human prothymosin α cDNA (Eschenfeldt and Berger, 1986) suggested that only one of the genes, designated PTMA, contained introns (Eschenfeldt *et al.*, 1989). Extensive sequencing of this gene and partial sequencing of the 5′ ends of the remaining genes (Eschenfeldt *et al.*, 1989) supported this conclusion and indicated, further, that all isolated human prothymosin α cDNAs (Eschenfeldt and Berger, 1986; Gomez-Marquez *et al.*, 1989; Goodall *et al.*, 1986) were derived from PTMA transcripts. In view of these findings, it was proposed that the remaining genes were likely to be nonfunctional, processed pseudogenes (Eschenfeldt *et al.*, 1989). This proposition was based, in part, on the observation that only the intron-containing gene possessed both CCAAT and TATA sequence elements in an upstream configuration similar to that of many other genes transcribed by RNA polymerase II (Breathnach and Chambon, 1981; Lewin, 1985).

Complete sequence analysis of four of the five putative processed pseudogenes (PTMAP1–4) and a more thorough examination of the fifth (PTMAP5) indicated that most aspects of our initial interpretation were correct. An alignment of the DNA sequences is presented in Fig. 1. To facilitate the comparison, the introns of PTMA (Eschenfeldt *et al.*, 1989) have been omitted, and deletions, duplications, and insertions of greater than 20 bp with respect to this gene have been indicated by symbols (Fig. 1A) and shown in detail (Fig. 1B). A diagram depicting the alignment and showing important sequence elements is presented as Fig. 2.

The PTMAP5 sequences detailed in Fig. 1A were contained in a plasmid clone isolated from a partial human genomic library made with *Eco*RI-restricted DNA (Eschenfeldt *et al.*, 1989); a T to C transition at base pair 695 created an *Eco*RI site, which truncated the gene and made it impossible to isolate it within the confines of one clone. We have made several attempts to recover the missing sequences, both by screening additional libraries with PTMAP5-specific probes and by amplifying the junction regions of ligated genomic circles using oligomers directed away from known sequences (reverse PCR; Silver and Keerikatte, 1989); all of these efforts have proven futile.

Genes PTMAP1–5 Are Processed Pseudogenes

Processed pseudogenes are DNA copies of mRNAs incorporated into the genome at sites other than those occupied by the parental genes (Vanin, 1985; Wilde, 1986). It is believed that these pseudogenes are inserted at staggered breaks in double-stranded DNA, but it is unclear whether the insertion events occur at the same time the DNA copies are made, or afterward (Wilde, 1986). Characteristically, processed pseudogenes exhibit four main structural features: (i) they lack introns; (ii) they possess poly(A) tail remnants adjacent to the last transcribed bases in mRNA; (iii) they contain short direct repeats of 5–25 bases immediately upstream and downstream of sequences found in mRNA; and (iv) they are flanked by sequences that are very different from those surrounding the transcribed regions of their parental genes (Vanin, 1985; Wilde, 1986). In addition, processed pseudogenes display a remarkable degree of sequence conservation; homologies of 90% or greater are frequently observed when processed genes and the relevant segments of their parental genes are compared (Vanin, 1985; Wilde, 1986).

Upon close examination of the sequences shown in Fig. 1A, it became clear that genes now designated PTMAP1–4 met all of the structural requirements for classification as processed pseudogenes. Furthermore, these genes exhibited ~85–95% sequence homology with respect to the gene. It is noteworthy that the direct repeats (singly underlined in Fig. 1A) flanking pseudogenes 1–4 are A/T-rich (average of 76% A and T) and that no clear line of demarcation can be defined to indicate where the vestigial poly(A) tails end and the 3′-end repeat sequences begin. The high A/T content of these flanking sequences is consistent with the hypothesis that processed pseudogenes are inserted predominantly at A/T-rich loci (Vanin, 1985; Wilde, 1986).

We believe that PTMAP5 is also a processed pseudogene. Although there is insufficient sequence information to make a definitive assignment, the following facts support this view: it contains none of the introns found in the functional gene; it is located at a site in the genome other than that occupied by the gene; and it has 5′-end sequences that are as A/T-rich as those adjacent to the 5′ ends of the other processed genes.

If processed pseudogenes are, in fact, DNA copies of mRNAs, they should contain only those residues found in mRNA molecules. Significantly, prothymosin α pseudogenes 1–4 contained nucleotides at their 5′ and 3′ ends that were not represented in our full-length cDNA. Each of these genes had 3–6 extra bases situated between its 5′ repeat sequences and the genomic transcriptional start site suggested by the cDNA (residue 1 in Fig. 1A); the four pseudogenes also had at least 3 additional bases

A

```
                                                                            *
         -60       -50        -40        -30        -20        -10          1         10         20         30
           |         |          |          |          |          |          |          |          |          |
➤PTMA       GGGAAGCCGA GCGCCGCCCA CTAATCTATA TTAAAGCTTC TGGCGCCGCG TGAGTCCCCA ACTGGCTGCT CTGAAAAGCC ATCTTTGCAT
PTMAP1      GaaAgattat agataGagaA gctgcCccTt TgggccCTct TtaaGaaGtG aGgacCCCCA ACTGGCTGCT CTGAAAAGCC ATCTTTGCAT
PTMAP2      cacttatgaA GCttaGtttg gctggagATg aaAttctagg Ttaaaaattc TtttattCCc ACTGGCTGCT CTGAAAAGCC ATCTTTGCAT
PTMAP3                                   ccc TgActcCTgt ataaaaaaat gacaTCaCCc ACTGGCaGCT CTGAAAAGCC ATCTTTGCAT
PTMAP4      aaacAaCtct tacCCtCCtc CaAATCTAag gacAtttgat gaaaaaCatt gtccTgCCCc ACTGGCTGCT CTGAAAAGCC gTCTTTGCAT
PTMAP5      aatgAaaaGA tgcCacattA aTtgTagAcA accAtttTaa gatatatttc attccgCCCc ACTGGCTGCT CTGAAAAGCC ATCTTTGCAT

                    40         50         60         70         80         90        100        110        120
                     |          |       ▼2 |          |          |          |          |          |          |
PTMA        TGTTCCTCAT CCGCCTCCTT GCCCGCCGCA GTCGCCTCCG CCGCGCGCCT CCTC-GCCGC CGCGGACTCC GGCAGCTTTA TCGCCAGAGT
PTMAP1      TGTTCCTggT tCGgtgtCcT GCTCaCCaCA GCCaCCTCCG CCatGCaCtT CCTCtGCtGC CtCaGAGTCt GGCAGCTTaA TCGaCAtAGT
PTMAP2      TGTTCCTCgT CCGCCTCCTT GCTCGCCGCA GCCGCCTCCG CCaCGCGCCT CCTC-GCCGC CGCGGACTCC GGCAGCTTTA TCGCCAGAGT
PTMAP3      TGTTCCTtgT CCGgCTCCTT GCTCGCCGgA GCCGCCTtta CCGCt///// ////////// /GCGGACTCC GGacaCTTcA TCaCCAcAGT
PTMAP4      TGTgCgTCgT CaGCCTCCTT GCTCGCCGCA GCCGCCTC// ////////// /////GCCGC CGCGGACTCC GGCAGCTTTA TCGCCAGAGT
PTMAP5      TGTT////// ///CCTCCTT GCTCGCCGCA GCCGCCTCCG CCGCGCGCCT CCTC/GCCGC CGCGGACTgC GGCAGCTTTA TCGCCAGAGT

                   130        140        150        160        170        180        190        200        210
                     |          |          |          |          |          |          |          |          |
PTMA        CCCTGAACTC TCGCTTTCTT TTTAATCCCC TGCATCGGAT CACCGGCGTG CCCCACCATG TCAGACGCAG CCGTAGACAC CAGCTCCGAA
PTMAP1      CCCcaAACTC TCaCTTTCTT cTTAATCCCt TGCATCGGAT CACCGctGTG CCCCACCATG TCAGAgGCAG ttGTgGACAC aAGCTCCGtg
PTMAP2      CCCTGAACTC TCGCTTTCTT TTTAATCgCC TGCATCGGAT CACCGGCGTG CCCCACCATG TCAGACGCAG CCGTAGACAC CAGCTCCGAA
PTMAP3      CCCTGAACTC TCGCTTTCTT TTTAATCCCC TGCATCGGAT CACtGGtGTG CCggACCATG TCAGACGCAG CCGTAGACAC CAGCTCCGAA
PTMAP4      CCCTGAACTC TCGCTTTCTT TTTtATCCCC TGCATCGcgT CACCGGCGTG CCCCACCATG TCAGACGCAG CCGTAGACAC CAGCTCCGAA
PTMAP5      CCCTGAACTC TCGCTTTCTT TTTAATCCCC TGCATCGGAT CACCGGCGTG CCCCACCATG TCAGACGCAG CCGTAGACgC CAGCTCCGAA

                   220        230        240        250        260        270        280        290        300
                     | ▼IVS1    |          |          |          |          |          |          |    ▼IVS2 |
PTMA        ATCACCACCA AGGACTTAAA GGAGAAGAAG GAAGTTGTGG AAGAGGCAGA AAATGGAAGA GACGCCCCTG CTAACGGGAA TGCTAATGAG
PTMAP1      ATCACCACCA AGGACTTcAA GGAGAAG/// ////TTGTGG AgGAGGCAGA AAgTGGAAGA GACGCCCaTG CTAACGGGAA cGCTAATGAG
PTMAP2      ATCACCACCA AGGACTTAAA GGAGAAGAAG GAAGTTGTGG AAGAGGCAGA AAATGGAAGA GACGCCCCTG CTAACGGGAA TGCTAATGAG
PTMAP3      ATCACCACCA AGGACTTAAA G///AAGAAG GAAGcTGTGG AgGAaGCgGA AAATGGAAGA GACaCCCCTG CTAAtGGGAA gGCTAATGAG
PTMAP4      ATCACCACCg AGGACTTAAA GGAGAAGAAG GAAGTTGTGG AAGAGGCgGA AAATGGAAGA GACGCCCCTG CTcACGGGAA TGCTAATGAG
PTMAP5      ATCACCAtCA AGGACTTAAA GGAGAAGAAG GAAGTTGTGG AAGAGGCAGA AAATGGAAGA GACGCCCCTG CTAACGGGAA TGCTAATGAG

                   310        320        330        340        350        360        370        380        390
                     |          |          |          |          |          |          |          |    IVS3▼ |
PTMA        GAAAATGGGG AGCAGGAGGC TGACAATGAG GTAGACGAAG AAGAGGAAGA AGGTGGGGAG GAAGAGGAGG AGGAAGAAGA AGGTGATGGT
PTMAP1      GAAAATGGGG AGCAGGAGGC TGACAAcGAG GTAGAtGAAG AAGAGGAAcA gGGTGGGGAG aAAGAGGAGa AGGAAGAgGA AGGTGATGGT
PTMAP2      GAAAATGGGG AGCAGGAGGC TGACAATGAG GTAGAtGAAG AAGAGGAAGA AGGTGGGGAG GAAGAGGAGG AGGAAGAA// /GGTGATGGT
PTMAP3      GAAAATGGGG AGCAGGAaGC TGACAATGAa GTAGAtGAAG AAGAGGAAGA AGGTcGGGAG GAAGAcGAcG AGGAAGAAGA AGGcGATGGT
PTMAP4      GAAAATGGGG AGCcGGA/// TGACAAcGAG GTAGAtGAAG AAGAGGAAGA AGGTGGGGAG GAAGAGGAGG AGGAAGAA// /GGTGATGGT
PTMAP5      GAAAATGGGG AGCAGGAGGC TGACAgTGAa GTAGAtGAAG AAGAGGAAGA AGGTGGGGAG GAAGAGGAGG AGGAAGAAGA AGGTGATGGT

                   400        410        420        430        440        450        460        470        480
                     |          |          |          |          |          |          | ▼IVS4    |          |
PTMA        GAGGAAGAGG ATGGAGATGA AGATGAGGAA GCTGAGTCAG CTACGGGCAA GCGGGCAGCT GAAGATGATG AGGATGACGA TGTCGATACC
PTMAP1      GAaGAAaAGa AcGGAGATGA AaAcGAaGcA GCTGAGgCgG /TAtGGaCAA atGGGCAGCT GAtGATGATG AaGATGACGA TGTtGATACC
PTMAP2      GAGGAAGAGG ATGGAGATGA AGATGAGGAA GCTGAGaCAG CTACGGGCAA GCGGGCAGCT GAAGATGATG AGGATGACGA TGTCGATACC
PTMAP3      GAGGAAGAGG ATGGtGATGA AGAcGAGGAA GtTGAGTCcG CTAgGt/CAA GCGGGCAGCT GAAGATGATG AGaATGAtGA TGcCtATACC
PTMAP4      GAGGAAGAGG AcGGAGATGA AGATGAGGgA GCTGAGTCAG CTACGGGCAA GCGGGCAGCT GAAGATGATG AGGATaACGA TGTCGATACC
PTMAP5      GAGGAAGAGG ATGGAGATGA AGATGAGGAA GCTGAGTCAc CTACGGGCAA GCGGGCAGCT GAAGATGATG AGGATGACGA TGTCGATACC

                   490        500        510        520        530        540        550        560        570
                     |          |          |          |          |          |          |          |          |
PTMA        AAGAAGCAGA AGACCGACGA GGATGACTAG ACAGCAAAAA AGGAAAAGTT AAACTAAAAA AAAAAAGGCC GCCGTGACCT ATTCACCCTC
PTMAP1      AAGcAGCAGA AGgCCagtGA GGATGAtTAG ACAGCAAAAA AaGAAAAGTT AAACTttAAA tt//AAGGCC aCCGTGACCT ATTCACCCTC
PTMAP2      AAGAAGCAGA AGACCGACGA GGATGACTAG ACAGCAAAAA AGGAAAAGTT AAACTAAAAA AAAA//GGCC GCCtTGACCT ATTCACCCTC
PTMAP3      AAGAAGCAGA AGACCaACaA GGATGACTAG ACAGCAAAAA AGGAAAtGTT Aggagg//// ////////// ///GTGACCT ATTCACCCTC
PTMAP4      cAGAAGCAGA AGACCGACGA GGATGACcAG ACAGCAAAAA AGGAAAAGTT AAACTAAAAA AAAAA/GGCC GCCGTGACCT ATTCACCCTC
PTMAP5      AAGAAGCAGA AGACCGACGA GGATGACTAG ACAGCAAAAA AGGAAAAGTT AAACTAAAAA AAAA//GGCC GCCGTGACCT ATTCACCCTC

                   580        590        600        610        620        630        640        650        660
                     |          |          |          |          | ▼3       |          |          |          |
PTMA        CACTTCCCGT CTCAGAATCT AAACGTGGTC ACCTTCGAGT AGAGAGGCCC -GCCCGCCCA CCGTGGGCAG TGCCACCCGC AGATGACACG
PTMAP1      CACTTCCCaT CTCAGAATCT AAACaTGGTt gCCcTCGAG/ AGgcctGCtt -GCCCtCC/A C///aGaCAG TGCCACT/GC AGATGACAgG
PTMAP2      CACTTCCCGT CTCAGAATCT AA/CGTGGTC ACCTTCGAGT AGAGAGGCCC -GCCCGCCCA CCGTGGaCAG TGCCACCCGC AGATGACACG
PTMAP3      CACTTCCtGT CTCAGAATCT AcAtGTGGTC ACCTTtGAGT AttGAGGCCC -GCCCGCCCA CCGaGGGCAa TGCCACCCGC AGATGACAtG
PTMAP4      CACTTCCCGT CTCAGAATCT AAACGTGGTC ACCTTCGAGT AGAGgGGCCC -GCCCGCCCA CCGTGGGCAG TGCCACCCGC AGATGACACG
PTMAP5      CACTTCCCGT CTCAGAATCT AAACGTGGTC ACCTTgaAGT AGAGAGGCCC gGCCtGCCCA CCGTGGGaAG TGCCtCCCaC AGATGAtACG
```

FIG. 1. (A) Alignment of human prothymosin α gene family sequences. The parental gene is designated by an arrow. For the purposes of alignment, the introns of this gene have been omitted; the position of each intron is indicated by an inverted triangle followed by the intron

```
                  670        680        690        700        710        720        730        740        750
                    |          |          |          |          |          |          |          |          |
PTMA       CGCTCTCCAC CACCCAACCC AAACCATGAG AATTTGCAAC AGGGGAGGAA AAAAGAACCA AAACTTCC-A AGGCCCTGCT TTTT----TT
PTMAP1     CaCTCaCCAC CACCCAACCC AAACCA/GAG AATTTGCAAC AGaGGAGGAA AAAAGAACCA AAACTTCC-A AGGtCtTGCT cTTT----//
PTMAP2     CGCTCTCCAC CACCCAACCC AAACCATGAG AATTTGCAAC AGGGGAGGAA AAAAGAACCA AAACTTCC-A AGGCCCTGCT TTTT----TT
PTMAP3     CGCTCTCCAC CACCCAACCC AAACCATGAG AATTTGCAAC AGGGGAGGAA AAAAGAACCA AAACTTCCcA AGGCCCTGCT TTTT----Tc
PTMAP4     CGCTCTCCAC CACCCAACCC AAACCATGAG AATTTGCAAC AGGGGAGGgA AAAAGAACCA AAACTTCC-A AGGCCC/GCT TTTTttttTT
PTMAP5     CGCTCTCCtC CACCCAACCC AAACCATGAG AATTc

                  760        770        780        790        800        810        820        830        840
                    |          |          |          |          |          |          |         ▼1a        |
PTMA       CTTAAAAGTA CTTTAAAAAG GAAATTTGTT TGTATTTTTT A-TTTACATT TTATATTTTT GTACATATTG TTAGGGTCAG CCATTTTTAA
PTMAP1     //TAAAAGTA CTTTAAAAAG GAAgTTTGTT TGTATTTTTT A-TTTACATT TTATATTTTT GTACATATTG TTAGGGTCAG tCATTTTTAA
PTMAP2     CTTAAAAGTA CTTTAAAAAG GAAATTTGTT TGTATTTTTT A-TTTACATT TTATATTTTT GTACATATTG TTAGaGTCAG CCATTTTTAA
PTMAP3     CTTAAAAaTA CTTTAAAAAG GAAgTTTGTT TGcATTTTTT AaTTTACATT TTATATTTTT GgACATATTG TTAGGGTCAG CCATTTTTAA
PTMAP4     CTTAAAAGTA CTTTAAAAAG GAAA////cT TGTATTTTTT A-TTTACATT TTATATTTTT GTACATATTG TTAGGGTCgG CCATTTTTAA
PTMAP5

                  850        860        870        880        890        900        910        920        930
                    |          |          |          |          |          |          |          |          |
PTMA       TGATCTCGGA TGGCCAAACC AGCCTTCGGA GCGTTCTCTG TCCTACTTCT GACTTTACTT GTGGTGTGAC -ATGTTCATT ATAATCTCAA
PTMAP1     TGATCTCaGA TGaCCAAgCC AGCCTTtGGA GgGTTCTCTG TCtTACTTCT GACTTTACTT GTGGTGTGAC cATaTTCATT ATAATCTCAA
PTMAP2     TGATCTCcGA TGaCCAAACC AGCCTTCGGA GCGTTCTCTG TCCTACTTCT GACTTTACTT GTGGTGTGAC cATGTTCATT ATAATCTCAA
PTMAP3     TGATCTCaGA TGaCCAAACC AGCCTTCaGA GCGTTCTCTG TCCTgCTTCT aACgTcACTT GTGaTGTGAC cATGTTCgTT ATAATCTCAA
PTMAP4     TGATCTCGGA TGaCCAAACC AGCCTTCGGA GCGTTCTCTG TCCTACTTCT cACTTTACTT GTGGTGTGgC cATGTTCATT ATAATCTCAA
PTMAP5

                  940        950        960        970        980        990       1000       1010       1020
                    |          |          |          |▼1b       |          |          |          |          |
PTMA       AGGAGAAAAA AAA------- -----CCTTGT --AAAAAAA GCAAAAATGA CAACAGAAAA ACAATCT-TA -TTCCGAGCA TTCCAGTAAC
PTMAP1     AGGAGGAAAA AAAAAAAAAA AAAAACCTTGT TTAAAAAAA aCAAAAAcaA CAACAacAAc AaAAagTcTc gTTCtGAGCA TTCCAGTAgC
PTMAP2     AGGAGAAAAA AAA------- -----CCTTGT --AAAAAAA GCAAAAATGA CAACAGAAAA ACAATCT-TA -TTCCGAGCA TTCCAGTAAC
PTMAP3     AGGAGAAAAA AA/------- -----CCTTGT --AAgAcAA GCAAAAAcGA CAACAGAAAA ACAATCT-TA -TTCtGAGCA TTCCAGTAAC
PTMAP4     AGGAGAAAAA AAAAA----- ------CTTGT --AAAAAAt GCAAAAATGA CAACAGAAAA ACcATCT-TA -TTCCGAGCA TTCCAGTAAC
PTMAP5

                 1030       1040       1050       1060       1070       1080       1090       1100       1110
                    |          |          |          |          |          |          |          |          |
PTMA       TTTTTT-GTG TAT--GTACT TAGCTGTACT ATAAGTAGTT GGTTTGTATG AGATGGTTAA AAAGGCCAAA GATAAAAGGT TTCTTTTTTT
PTMAP1     TTcTTTaGTG TAT--GTAgT TAGtTGTACc ATAAGTAGTT GGTTTGTgTG AGATGGTTAA AAAGGCCAAA GATAAAtGtT TcaTTTaTTT
PTMAP2     TTTTTT-GTG TAT--GTACT TAGCTGTACT ATAAGTAGTT GGTTTGTATG AGATGGTTAA AAAGGCCAAA GATAAAAGGT TTCTTTTTTT
PTMAP3     TTTTTT-GTG TgTgcGTACT TAGCTGTACT ATAAGTAGTT GGTTTGTATG AGATGGTTAA AAgGGCCAAA GATAAAAGGT TTCTTTTTTT
PTMAP4     TTTTTT-GTG TAT--GTACT TAGCTGTACT ATAAGTAGTT GGTTTGTATG AGATGGTTAA AAAGGCCAAA GATAAAAGGT TTCTTTTTTT
PTMAP5

                 1120       1130       1140       1150       1160       1170       1180       1190       1200
                    |          |          |          |          |          |          |          |          |
PTMA       TTCCTTTTTT GTCTATGAAG TTGCTGTTTA TTT------- --TTTTTGGC CTGTTTGATG TATGTGTGAA ACAATGTTGT CCAACAATAA
PTMAP1     /gCCTTTTTT GTCTATGAAa TgGCTGcTTA TTT------- --aTTTaGGC CTaTTTGATG TATGTGTGAA ACAATaTTGT gCAACAATAA
PTMAP2     T/CCTTTTTT GTCTATGAAG TTGCTGTTTA TTT------- --TTTTTGGC CTGTTTGATG TATGTGTGAA ACAATGTTGT CCAACAATAA
PTMAP3     T/CCTTTTcT GTCTATGAAG TTGCTGTTTA TTTatttatt taTTTTTtGC CTaTTTGAgG TATGTGTGAA ACAATGTTGT CCAACAATAA
PTMAP4     T/CCTTTTTT GTCTATGAAG TTGCTGTTTA TTT-----tt atTTTTTGGC CTGTTTGATG TATGTGTGAA ACAATGTTGT CCAACAATAA
PTMAP5

                                     *
                 1210       1220       1230       1240       1250       1260       1270       1280       1290
                    |          |          |          |          |          |          |          |          |
PTMA       ACAGGAATTT TATTTTGCTG AGTTGTTCTA ACAAAGCTGT CTCAAGCCTG GTTTTTCTGT TTCAGTTTCT TCAGACCTTC CAGGGCACAA
PTMAP1     ACccaAATTT TATTTTGCTG AGTTGTTCTA ACAgcaacaa aaagAagtTa aggaagagaa gaagaccagc aaAtgCaacC acaGagtgAc
PTMAP2     ACAGGAATTT TATTTTGCTG AGTTGTTCTA ACAAAaaaaa aaattcttTt aTTTaagaaT gTtgaaTatT ggccccCact CtcttCtggc
PTMAP3     ACAGGAATTT TATTTTcCTG AGTTGTTCTA ACAAcaacaa aaatgaCaTc agcTgggcGc ggtAGcTcaT gCctgtaaTc acaGcacttt
PTMAP4     ACAGGAATTT TATTTTGCTG AGTTGTTCTA gCAAAaaaaa aaaAgaaaaa aaaaagaaaa acattgTTCT gatGAaaaTC acttGgAatg
PTMAP5
```

number (e.g., intron 1 = ▼IVS1). Asterisks (*) define the first and last transcribed bases in our full-length cDNA (Eschenfeldt and Berger, 1986); the gene sequence lying between these asterisks reflects that present in most prothymosin α mRNAs (Manrow and Berger, unpublished). Lowercase letters designate nucleotide differences with respect to the gene. All deletions, insertions, and duplication of less than 20 bases are shown in A; larger duplications and insertions are detailed in B. The positions of the larger duplications and insertions are indicated in A by inverted triangles and numbers identifying the relevant genes (e.g., ▼2 refers to PTMAP2). Where necessary, dashes (–) have been introduced to preserve the alignment, thereby accommodating insertions and duplications; deleted bases are indicated by slash marks (\). Repeated sequences within a given gene are singly underlined; potential translational initiation and termination codons are doubly underlined. Coding region *AccI* sites at positions 193 and 331 are italicized. (B) Large duplications and insertions in prothymosin α gene family members. Parts 1 and 2: The sequences of *Alu* elements in PTMAP1. The symbol "a" refers to the element situated between bases 829 and 830, and "b" refers to the element inserted between bases 971 and 972. Parts 3 and 4: Sequences of large duplications in PTMAP2 and 3. The duplicated material consists of two additional copies of sequences found once in the gene; the arrows define the limits of each duplication unit. The underlined bases in PTMAP2 duplications denote residues found at the 3′ end of the parental sequence; the underlined bases in the PTMAP3 units represent nucleotides present at the 5′ end of the parental sequence (see text).

B

1. PTMAP1: *Alu* element "a"

```
        10         20         30         40         50         60         70         80         90
         |          |          |          |          |          |          |          |          |
TTTTTTTTTT CTTTGAGACG GAGTCTAGCT CTGTCGCCAG GCTCAAGTGC AGTGGTGCGA TCTTGGCTCA CCGCAAGCTC CACCTCCTGG

       100        110        120        130        140        150        160        170        180
         |          |          |          |          |          |          |          |          |
GTTCAAGTGA TTCTCCTGCC TCAGCCTCCT GAGTAGCGGG GATTACAGGC GCCCGCCACC ACACCCAGCT AATTTTTGTA TTTTTAGCAG

       190        200        210        220        230        240        250        260        270
         |          |          |          |          |          |          |          |          |
AGACAGGCTT TCACCAGGTT GGCCAGGATG GTTTCTATCT CCTGACCTTG TGATCCACCT ACCTCGGCCT CCCAAAGTGC TCGAATTACA

       280        290        300
         |          |          |
GGCGTGAGCA CCGGCGCCAG GTT
```

2. PTMAP1: *Alu* element "b"

```
        10         20         30         40         50         60         70         80         90
         |          |          |          |          |          |          |          |          |
AAAAAAAGCC TGGGCGCGGT GGCTCGCGTG TAATCCCAGC ACTTTGGGAG GCCGAGGTGG GTGGATCACG AGGTCAGAAG ATCGAGACCA

       100        110        120        130        140        150        160        170        180
         |          |          |          |          |          |          |          |          |
TCCTGGCTAA CATGGTGAAA CCCCCTGTCT ACTAAAAATA CAAAAAATTA GCCAGGCGTG GTGGCGGGAG CCTGTAGTCC CAGCTACTTG

       190        200        210        220        230        240        250        260        270
         |          |          |          |          |          |          |          |          |
GGAGGCTGAG GCAGGAGAAT GGCGTGAACC CGGGAGGCAG AGCTTGCAGT GAGCCAAGAT TTGCCACTGC ACTCCAGCCT GGGCAACAGA

       280        290
         |          |
GCGAGACTAC ATCT
```

3. PTMAP2 duplication:

```
        10         20         30         40
↓        |         ↓|          |      ↓   |
GGCAGCCTCC TTGCTCGCGG CAGCCTCCTT GCTCGC
```

4. PTMAP3 duplication:

```
        10         20         30         40         50         60         70         80         90
    ↓    |          |          |          |      ↓   |          |          |          |         ↓|
TAACCCACCC ACTGCGGGCA GTGCCACCCG CAGATGACAC GGCCCACCCG CCCACCGAGG GCAGTGCCAC CCGCAGATGA CACGGCCCAC
```

FIG. 1—*Continued*

lying between the last transcribed nucleotide in the cDNA (i.e., base 1230) (Eschenfeldt and Berger, 1986) and the beginning of their poly(A) tail sequences. Comparison of the gene and the processed pseudogenes revealed that these extra nucleotides were virtually identical to similarly positioned residues in the parental gene. From this finding, we infer that the pseudogenes were derived from transcripts initiated at several sites 5′ to the site represented in the cDNA, and that these transcripts were cleaved for polyadenylation at a different downstream site(s) as well. The homology between PTMAP5 and the gene also extended beyond the limits defined by the cDNA (see residues −1 to −4 in Fig. 1A), suggesting that the mRNA used to generate this processed gene was also initiated at an upstream site. Multiple transcriptional initiation sites have been reported (e.g., Shelness and Williams, 1984), and heterogeneity in polyadenylation cleavage site selection has been observed (e.g., Sasavage *et al.,* 1982). Since the creation of each processed pseudogene is believed to occur independently (Vanin, 1985; Wilde, 1986), the presence of extra bases in all of the prothymosin *α* processed genes strongly suggests that upstream transcriptional start sites and the downstream polyadenylation cleavage site(s) are used more frequently than those defined by our cDNA. It appears, therefore, that many prothymosin *α* transcripts are initiated with C residues. While the majority of eukaryotic pre-mRNAs are initiated at purines embedded in pyrimidine-rich tracts (Breathnach and Chambon, 1981), pyrimidine starts are encountered (e.g., Dudow and Perry, 1984). Thus, our cDNA, with its 5′-terminal adenosine residue, is probably full-length,

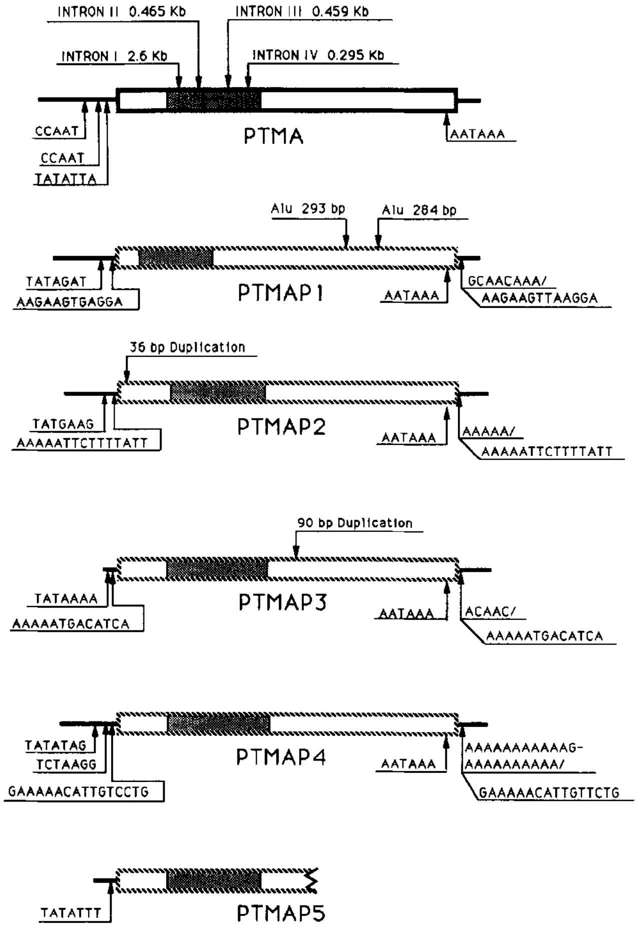

FIG. 2. The human prothymosin α gene family. The positions of important structural features are shown. The rectangular boxes denote regions of each gene corresponding to sequences found in prothymosin α mRNAs; the shaded areas indicate open reading frames. The 5'- and 3'-flanking regions of each gene are represented by bold lines. In the 5'-flanking regions, the relative positions of CCAAT, TATA, and 5'-end repeat sequences are shown. In the boxed regions, the positions of introns, large duplications, large insertions, and polyadenylation cleavage signals are indicated. In the 3'-flanking regions, residual poly(A) tail sequences and 3'-end repeat sequences are shown; the slash marks (\) delineate arbitrary boundaries between where poly(A) tails end and 3' end repeats commence.

reflecting one of a multiplicity of transcriptional start sites in the prothymosin α gene (see Fig. 1A).

Several of the processed pseudogenes contained sizable duplications and/or insertions not found in the gene; the most significant of these are detailed in Fig. 1B. PTMAP1 contained two *Alu* elements (Schmid and Jelinek, 1982) that are separated from one another by approximately 140 bases (also see Fig. 1A). Element "a" is located between residues 829 and 830 and lies in the 3' → 5' orientation (Schmid and Jelinek, 1982), whereas element "b" is situated between bases 971 and 972 in the 5' → 3' orientation.

Both PTMAP2 and 3 contained three copies of sequences that are found only once in the gene. PTMAP2 had three adjacent copies of the sequence 5'-GCCTCCTTGC(C/T)CGC(C/G)GCA-3', representing bases 43–60 of the gene. The optimized alignment shown in Fig. 1A dictated that the extra copies of this sequence be positioned inside the parental sequence (between bases 56 and 57), indicating that sequence rearrangement occurred during the duplication process. The PTMAP2 sequences detailed in Fig. 1B reflect this juxtaposition; the last four bases of the parental sequence (underlined) appear at the 5' end of each duplication unit. PTMAP3 contained three copies of residues 617–660 of the gene (see Figs. 1A and B). The two adjacent, extra copies also appear to lie within the parental sequence between bases 621 and 622 (Fig. 1A), a location consistent with reorganization of the gene. Thus, in Fig. 1B, the first five nucleotides of the parental sequence (underlined) appear as the last five bases of the duplication units.

Transcription of Prothymosin α Genes

Regulatory elements similar to those located near the 5' ends of many expressed genes (Breathnach and Chambon, 1981; Lewin, 1985) were found upstream of the functional prothymosin α gene. The TATA element of this gene, which extends from bases −27 to −33, is flanked by G/C-rich sequences (see Figs. 1A and 3). Two CCAAT elements spanning the regions from −73 to −78 and −106 to −110 (see Eschenfeldt *et al.,* 1989) were also identified; the CCAAT sequence at −78 is separated from potential transcriptional start sites by the consensus distance interval of 70–80 bp (Breathnach and Chambon, 1981; Lewin, 1985). It should be noted that CCAAT elements are not obligatory for gene expression, but when present, can be involved in both positive and negative transcriptional regulation (Barberis *et al.,* 1987).

To date, convincing evidence for expression of a processed pseudogene has not been forthcoming (Vanin, 1985). Almost all of the known genes either lack transcriptional regulatory elements or contain multiple lesions that prevent them from yielding proteins resembling the parental gene products (Vanin, 1985). It was extremely surprising, therefore, to discover that most of the prothymosin α processed genes possessed both fairly intact reading frames (see below) and TATA-like se-

	SEQUENCE							5' START POSITION[a]
CONSENSUS[b]	T_{82}	A_{97}	T_{93}	A_{85}	A_{63} T_{37}	A_{83}	A_{50} T_{37}	-25 to -30
PTMA	T	A	T	A	T	T	A	-33
PTMAP1	T	A	T	A	G	A	T	-53
PTMAP2	T	A	T	G	A	A	G	-56
PTMAP3	T	A	T	A	A	A	A	-21
PTMAP4-1	T	A	T	A	T	A	G	-160
PTMAP4-2	T	C	T	A	A	G	G	-36
PTMAP5	T	A	T	A	T	T	T	-18

FIG. 3. Human prothymosin α gene family TATA elements. [a]5' start position refers to the number of the first nucleotide of the element. [b]Breathnach and Chambon (1981) and Lewin (1985).

	-5					+1			+4
CONSENSUS[a]	C	C	A G	C	C	A	U	G	G
PTMA	C	C	A	C	C	A	U	G	U
PTMAP1-1	C	C	G	C	C	A	U	G	C
PTMAP2-2	C	C	A	C	C	A	U	G	U
PTMAP2	C	C	A	C	C	A	U	G	U
PTMAP3	G	G	A	C	C	A	U	G	U
PTMAP4	C	C	A	C	C	A	U	G	U
PTMAP5	C	C	A	C	C	A	U	G	U

FIG. 4. Human prothymosin α gene family translational start site sequences. Two PTMAP1 translational start sites are indicated; PTMAP-1 refers to the first acceptable AUG codon encountered on putative transcripts; PTMAP-2 represents the prothymosin α initiator AUG codon. [a]Kozak (1987).

quence elements situated near their 5′ ends; two TATA-like elements were found adjacent to PTMAP4 (see Fig. 3). Only one of the potential TATA elements, that adjacent to PTMAP1, violated the consensus limits with regard to sequence; this element has a G residue in the fifth position, whereas the consensus rules dictate a near absolute requirement for an A or a T. G and C residues are tolerated at other positions in functional TATA elements, but such substitutions are rare (Breathnach and Chambon, 1981; Lewin, 1985). In view of these findings, the possibility of finding a functional prothymosin α processed pseudogene, other than PTMAP1, had to be seriously considered.

The sequences surrounding the putative TATA elements were analyzed. These flanking sequences are less G/C-rich than those surrounding the genuine TATA element of PTMA. The G/C content of the 20 bases preceding and the 20 bases succeeding the PTMA TATA sequence is 68%; the G/C content of a similar range of residues in the processed genes averages only 45%. Since transcriptional initiation usually occurs 20–30 bases downstream of TATA elements and since the polyadenylation cleavage signal of the gene (residues 1196–1201 in Fig. 1A) appears to be preserved in the processed genes, transcripts originating from most of these genes would be nearly identical in size to genuine prothymosin α mRNAs. PTMAP4 is a possible exception—the TATA-like element positioned at bases −154 to −160 has a sequence that is much closer to the consensus sequence than the element located at −30 to −36; use of this upstream element would yield transcripts approximately 100 bases longer than normal prothymosin α mRNAs, but not different enough in size to be distinguished electrophoretically.

Translation of Prothymosin α Genes

The translational start site sequences were examined for their adherence to consensus sequences that enhance translation (Kozak, 1986, 1987). As shown in Fig. 4, the upstream nucleotides (−1 to −5) of the gene (PTMA) are in compliance, but the nucleotide at +4 does not agree with the consensus choice. Nonetheless, prothymosin α mRNA is well translated *in vivo* and *in vitro* (Manrow *et al.,* 1991). The pseudogenes 2, 4, and 5 are identical with the gene at this location and if transcribed would, presumably, be translated as well as the known mRNA. In contrast, the initiator codon of PTMAP3 differs from the consensus motif. However, the presence of G, rather than C, residues at positions −4 and −5 (Fig. 4) would probably not abrogate the translation of these RNAs (Kozak, 1986). PTMAP1 raises a slightly different problem. Two adjacent nucleotide substitutions at bases 73 and 74 introduce an upstream ATG codon that appears to be in context for efficient translation, except perhaps for the problematic nucleotide at the +4 position. For this gene, the initiator codon used by the functional gene would now be an internal methionine codon. Although this downstream ATG codon is embedded in sequences identical with those of PTMA and three of the pseudogenes, there is no apparent reason for the translational machinery to favor it as a start site.

Figure 5 shows that most of the processed pseudogenes encode polypeptides that are quite similar to prothymosin α. The PTMAP2 protein, the most conserved, differs from prothymosin α at two positions; one of eight consecutive glutamic acid residues is deleted and a serine codon in the gene is a threonine codon in the pseudogene. The product of PTMAP5 is identical in size with the gene product but has four amino acid substitutions, which do not, however, alter the net charge. How these differences might affect prothymosin α function is unknown. Recently, we have shown that prothymosin α is a nuclear protein and that the KKQK amino acid motif at its carboxyl terminus constitutes part of a potent nuclear localization signal (Manrow *et al.,* 1991). The polypeptides encoded by PTMAP2 and 5 retain this targeting signal.

The proteins specified by PTMAP3 and 4 are not nearly as conserved and differ from authentic prothymosin α both in net negative charge and in size. The product of PTMAP4, regardless of the location of the transcriptional start site, contains seven amino acid substitutions relative to prothymosin α and an additional nine amino acids at the carboxyl terminus. One of the substitutions in this protein alters the nuclear targeting signal (KKQK to QKQK). However, the extra C-terminal amino acids include KKEK, a motif that might represent a new targeting signal or a suppressor of the effects of the aforementioned substitution. Acidic residues flanked by basic amino acids are found in the nuclear targeting signal of the *Xenopus* N1 protein (Kleinschmidt and Seiter, 1988), in one component of the nuclear localization signal of c-*myc* (Dang and Lee, 1988), and in the proposed nuclear targeting signal of the human estrogen receptor (Guiochon-Mantel *et al.,* 1989). PTMAP3 encodes a protein that is still more highly divergent; it is similar to prothymosin α from amino acid 1 to 85 (8 substitutions), but is completely

```
PTMAP1          mhflcclrvwqlnrhspqtltfflipcigsplcpt....

                        10        20        30        40        50        60
                         |         |         |         |         |         |
➧PTMA           MSDAAVDTSSEITTKDLKEKKEVVEEAENGRDAPANGNANEENGEQEADNEVDEEEEEGG
PTMAP1       ...MSeAvVDTSSvITTKDfKEKlwrrqkvEetpmltgtlmrkmgsrrlttr
PTMAP2          MSDAAVDTSSEITTKDLKEKKEVVEEAENGRDAPANGNANEENGEQEADNEVDEEEEEGG
PTMAP3          MSDAAVDTSSEITTKDLK-KKEaVEEAENGRDtPANGkANEENGEQEADNEVDEEEEEGr
PTMAP4          MSDAAVDTSSEITTeDLKEKKEVVEEAENGRDAPAhGNANEENGEp-dDNEVDEEEEEGG
PTMAP5          MSDAAVDaSSEITiKDLKEKKEVVEEAENGRDAPANGNANEENGEQEADsEVDEEEEEGG

                        70        80        90       100       110       120
                         |         |         |         |         |         |
PTMA            EEEEEEEEGDGEEEDGDEDEEAESATGKRAAEDDEDDDVDTKKQKTDEDD
PTMAP1
PTMAP2          EEEEEEE-GDGEEEDGDEDEEAEtATGKRAAEDDEDDDVDTKKQKTDEDD
PTMAP3          EEddEEEEGDGEEEDGDEDEEvESArssgqlkmmrmmmpiprsrrptrmtrqqkrkc
PTMAP4          EEEEEEE-GDGEEEDGDEDEgAESATGKRAAEDDEDnDVDTqKQKTDEDDqtakkekln
PTMAP5          EEEEEEEEGDGEEEDGDEDEEAESpTGKRAAEDDEDDDVDTKKQKTDEDD
```

FIG. 5. Polypeptides encoded by human prothymosin α gene family members. The arrow indicates the sequence of authentic prothymosin α, the product of the parental gene. Amino acid differences from the prothymosin α sequence are denoted by lowercase letters; deletions are indicated by dashes (–).

different at the carboxyl terminus. A single nucleotide deletion at base number 437 (see Fig. 1A) alters the reading frame, yielding a polypeptide that is seven amino acids longer than prothymosin α. The PTMA-like nuclear targeting signal is absent, but instead, this protein possesses carboxyl terminal PRSRRPTR and TRQQKRK, sequence motifs not unlike those found in the human HTLV I p27^x and polyoma large T-antigen nuclear targeting signals (SQRKRPPTP and VSRK RPP, respectively) (Siomi *et al.,* 1988; Richardson *et al.,* 1986).

The product of PTMAP1 is very different from prothymosin α. The upstream ATG codon is in frame with the ATG triplet (position 178) used by the functional gene. The new initiation codon introduces 35 amino acids on the amino-terminal side of the first 20 amino acids of prothymosin α. Thereafter, a 7-bp deletion (nucleotides 238–244; see Fig. 1A) causes a frameshift. Thus, PTMAP1 is the most highly divergent of the pseudogenes.

Search for Pseudogene Transcripts in Tissues

Because the pseudogenes retained signals for transcription and translation, the possibility of pseudogenomic mRNAs and proteins had to be considered. Evidence for expression of these genes was sought in seven normal human tissues—placenta, liver, thyroid, striated muscle, colon, stomach, and kidney—and one abnormal tissue—ovarian carcinoma. As a first step, coding-region-specific oligomers PsG up and PsG down targeted immediately upstream of the conserved genomic methionine initiator codon and immediately upstream of the genomic stop codon, respectively, were synthesized. These were tested in the polymerase chain reaction with cloned cDNA and pseudogenomic templates. As shown in Fig. 6A (lanes 1–6), amplification of the prothymosin α cDNA and pseudogenes 1–5 resulted in fragments of ~370 bp in every case. This band reflects amplification of the genes between positions 152 and 507 by oligomers each bearing nine unkeyed bases at their 5′ ends. The products were also produced in roughly equivalent amounts from nanogram quantities of each template. This point is important because the oligomers were targeted toward regions that are highly conserved among the gene family, but nevertheless, not identical; internal oligomer mismatches might have discriminated against three of the pseudogenes in the polymerase chain reaction.

Substrates for PCR were generated from RNA samples isolated from the eight tissues noted above; single-stranded cDNA was synthesized using a mixture of PsG down and oligo(dT) as the primers. A cursory inspection of the gene family indicated that all members contain an A-rich region in the vicinity of positions 516–546 that might serve as the initiation site for oligo(dT)-primed cDNA synthesis, in preference to the more distantly located poly(A) tail. However, PTMA is more richly endowed with A residues and more likely to be reverse transcribed than, for example, PTMAP3, which lacks 10 consecutive A's. Therefore, the PsG down primer was included in an attempt to equalize the probability of obtaining cDNA from all prothymosin α mRNAs in the samples. The cDNAs were then subjected to amplification in the presence of both PsG up and PsG down and analyzed. Figure 6B shows that all RNA samples gave rise to a prominent band at ~370 bp (lanes a–h), the same size as the products derived from the cloned templates (see Fig. 6D, lanes 4 and 5 for a side by side comparison). It can also be seen in Fig. 6B that treatment with pancreatic ribonuclease A before reverse transcription eliminated all amplification products (lanes i–p) and that no fragments were generated in the absence of template DNA (lane –) or in the presence of the ribonuclease sample used to degrade RNA (lane r). Taken together, the results indicate that all RNA samples con-

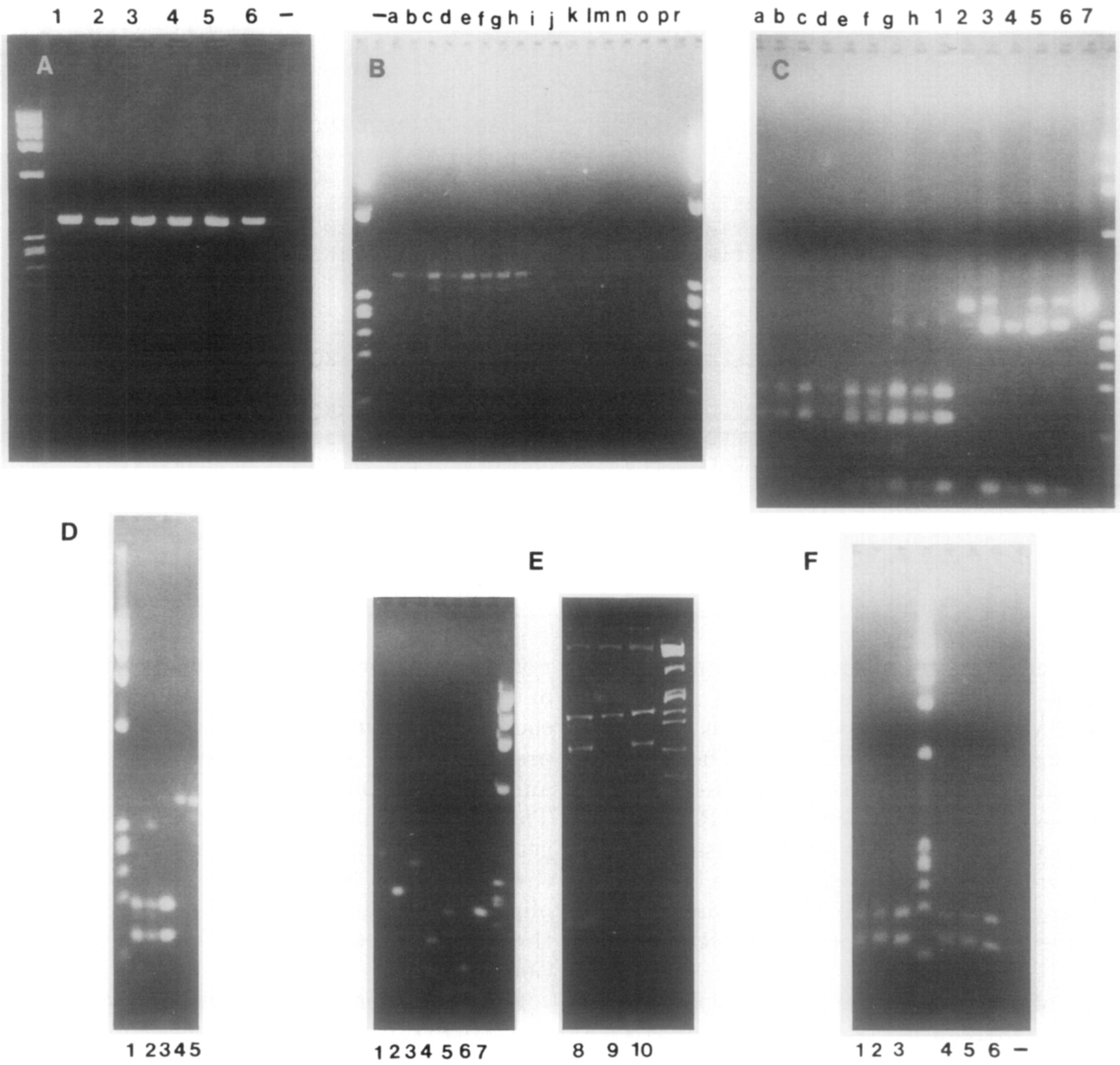
1 2 3 4 5 6 –
A
–a b c d e f g h i j k l m n o p r
B
a b c d e f g h 1 2 3 4 5 6 7
C
D
1 2 3 4 5
E
1 2 3 4 5 6 7
8 9 10
F
1 2 3 4 5 6 –

FIG. 6. Characterization of cloned human prothymosin α genes and tissue-derived single-stranded human cDNAs by means of the polymerase chain reaction. (**A**) Amplification of the cDNA clone derived from functional prothymosin α mRNA and of cloned pseudogenomic DNA templates as follows: lane **1**, PTMA cDNA; lane **2**, PTMAP1; lane **3**, PTMAP2; lane **4**, PTMAP3; lane **5**, PTMAP4; lane **6**, PTMAP5; and lane **–**, amplification in the absence of added DNA. Approximately 5% of the PCR products from each sample was analyzed electrophoretically in a 2.5% agarose gel. (**B**) Amplification of single-stranded cDNAs derived from RNA obtained from human tissues as follows: **a**, placenta; **b**, liver; **c**, thyroid; **d**, striated muscle; **e**, colon; **f**, ovarian carcinoma; **g**, stomach; and **h**, kidney. Samples **i–p** are identical to **a–h**, respectively, except that ribonuclease was added prior to the first-strand cDNA synthesis. In **–**, DNA was omitted and in **r**, 75 ng of the ribonuclease itself was tested with PCR. Electrophoretic analysis was carried out with 2.5% of the products of the polymerase chain reaction using a 2.5% agarose gel. (**C**) Analysis of the PCR products of genes and cDNAs after treatment with *Acc*I. Approximately 4 and 7% of the DNA produced by PCR from clones and single-stranded cDNAs, respectively, were cleaved with 15 units of *Acc*I for 1 h and 10 min in a volume of 30 μl. The starting material for lanes **1–6** was the DNA shown in lanes **1–6**, respectively, of **A**. The starting material for lanes **a–h** was the DNA shown in lanes **a–h**, respectively, of **B**. Lane 7 contains uncut PCR products from PTMA cDNA. Electrophoretic analysis was carried out in a 2.5% agarose gel. (**D**) Exhaustive digestion of the PCR DNA from stomach and kidney. PCR DNA fragments from stomach (**g** in part **B**), lane **1**; kidney (**h** in part **B**), lane **2**; and the PTMA cDNA (**1** in part **A**), lane **3** were cleaved overnight with 30 units of *Acc*I in submerged tubes under conditions otherwise identical with those in **C**. Lane **4** contains uncut PCR products from PTMA cDNA, and lane **5** contains uncut PCR products from stomach, for comparison. (**E**) Search for the putative pseudogenomic PCR products generated from kidney. The remaining PCR DNA from kidney (**h** in part **B**) was purified by spun G-50 Sephadex column chromatography, divided into four tubes containing, respectively, 12, 18, 23, and 46% of the total volume, and dried. In ascending order of quantity, the samples were cleaved with *Rma*I, lane **1**; *Hpa*II, lane **3**; *Bsm*AI, lane **5**; and *Bsa*HI, lane **8**, each in a total volume of 25 μl. An amount of enzyme equal to 10 to 25 units was used depending on the sample. As controls, 2% of the PCR products derived from PTMAP3 were digested with *Rma*I, lane **2**; from PTMAP4 with *Hpa*II, lane **4**; from PTMAP2 with *Bsm*AI, lane **6**; from PTMA with *Bsm*AI, lane **7**; from PTMAP 5 with *Bsa*HI, lane **9**; and from PTMA with *Bsa*HI, lane **10**. The samples cleaved with *Bsa*HI were analyzed electrophoretically in a 20% polyacrylamide gel; all others were resolved in 2.5% agarose. (**F**) Analysis of PCR products derived from single-stranded cDNAs from stomach (lanes **1–3**) and kidney (lanes **4–6**) by cleavage with *Acc*I. Each cDNA sample was subjected to the polymerase chain reaction in three independent experiments. An amount of DNA equal to 7% of the PCR products was cleaved overnight with *Acc*I as in **D** (lanes **1**, **2**, **4**, and **5**). In lanes **3** and **6**, the entire yield of PCR products was subjected to electrophoresis, the $\sim$370-bp band was excised from the gel, DNA was extracted and precipitated, and >30% of the recovered material was cleaved with *Acc*I and analyzed as in **D**. In each agarose gel, 1 μg of ϕX RF DNA digested with *Hae*III provided a series of markers; for the polyacrylamide gel, 1.5 μg of markers was resolved (unlabeled lanes).

tain prothymosin α transcripts, and that production of the PCR fragments is completely dependent upon prior reverse transcription. Apparently, the RNA samples did not contain prothymosin α genomic DNA sequences as contaminants.

The nature of PCR products was investigated by restriction endonuclease digestion. Whereas the PCR product of the cDNA has two *Acc*I sites (located at positions 193 and 331 in PTMA) and the equivalent product from PTMAP1 has none, fragments generated from any of the other four pseudogenes (PTMAP2–5) contain a single *Acc*I site. Accordingly, bands of 184, 138, and 52 bp are diagnostic for the known cDNA, 374 bp for PTMAP1, and 322 and 52 bp for any of the other pseudogenes (PTMAP 2–5). Figure 6C (lanes 1–6) details the results obtained when the PCR products of the cDNA and the five pseudogenes (PTMAP1–5) were cleaved with *Acc*I; although the cleavage was not always complete, the expected bands were obtained. Lanes a–h show the fragments generated from the PCR products arising from the eight RNA samples. In every case, fragments indicative of PTMA mRNA were visualized, but in two cases, the samples from stomach and kidney, a fragment of ~320 bp was observed as well. Since such a band might arise either from a genuine pseudogenomic transcript in the RNA samples or from incomplete cleavage with *Acc*I, the two suspicious samples were cleaved overnight with 30 units of *Acc*I to drive the reaction to completion. The results are shown in Fig. 6D. The suspicious band in the sample from stomach all but disappeared, whereas the ~320-bp band in the kidney sample clearly remained. For comparison, lane 3 (Fig. 6D) exhibits the PCR products of a larger amount of PTMA cDNA cleaved in an identical manner. The survival of the band at ~320 bp suggested several possibilities: the band might reflect a genuine pseudogenomic transcript of PTMAP2, 3, 4, or 5; it might indicate genomic contamination in either the reaction vessel or a pipet tip used to produce the kidney sample; or it might arise from an error introduced by the *Taq* polymerase in the downstream *Acc*I site of PTMA (position 331).

The first two possibilities were explored further by cleavage of PCR products with a series of restriction enzymes, each of which produced a fragment unique to a pseudogene under consideration. The four enzymes *Rma*I, *Hpa*II, *Bsm*AI, and *Bsa*HI, which distinguish PTMAP3, 4, 2, and 5, respectively, were initially used to digest the PCR products of the four aforementioned pseudogenes and the cDNA (PTMA); the expected bands appeared upon gel electrophoresis (Fig. 6E, lanes 2, 4, 6, 7, 9, and 10, and data not shown). Thus, cleavage of the ~370-bp PCR fragment of PTMAP3 (but not pseudogene 2, 4, or 5) with *Rma*I yielded a band of **~280 bp** (Fig. 6E, lane 2), and cleavage of the equivalent material from PTMAP4 with *Hpa*II generated a specific **~200-bp** band (lane 4). The PCR products of the cDNA remained intact upon treatment with the former enzyme, whereas the latter cleaved only once in the upstream oligomer, removing 22 bp (data not shown). The patterns obtained with the other two enzymes were more complex: *Bsm*AI cleaves the PCR products of the cDNA once, with resultant bands of ~250 and ~120 bp (Fig. 6E, lane 7). With PCR products of PTMAP2 only, the larger of these is digested further by *Bsm*AI to yield a specific band of **~160 bp** (Fig. 6E, lane 6). Similarly, *Bsa*HI cleaves the products of the cDNA and PTMAP2 and 4 (but not PTMAP3) only one time, generating bands of ~240 and ~130 bp (Fig. 6E, lane 10), but in this case it is the smaller of the two that is cleaved further to yield a fragment of **~80 bp** specific for PTMAP5 (Fig. 6E, lane 9).

With the establishment of specific bands (noted in boldface above) to identify each pseudogene in question, the kidney PCR products could then be analyzed more thoroughly. To ensure an equal probability of visualizing a pseudogene-specific band regardless of its size, the amount of material digested was inversely proportional to the size of the unique fragment being sought. Thus, almost fourfold more material was used in the search for PTMAP5, where the largest specific band is ~80 bp, than was used for PTMAP3, where the largest band is ~280 bp. As shown in Fig. 6E (lanes 1, 3, 5, and 8), the products of the kidney sample matched the products of the cDNA, with no detectable pseudogene-specific bands. The data suggested that neither a genomic contaminant nor a genuine pseudogenomic transcript was responsible for the suspicious 322-bp band.

The remaining possibility, that the band arose as a PCR artifact, seemed all too likely. Hence, the stomach and kidney single-stranded cDNAs were subjected to PCR three additional times, with the production of the expected ~370-bp band in every case (data not shown). One sample from each tissue was gel purified, before all samples were cleaved with *Acc*I. As detailed in Fig. 6G, neither the stomach (lanes 1–3) nor the kidney (lanes 4–6) contained mRNA derived from pseudogenes. Even the gel-purified samples (lanes 3 and 6) gave rise only to bands specific for the cDNA from the functional gene. The conclusion was that PCR had either destroyed the unique downstream *Acc*I site (position 331) or, perhaps, generated an inhibitor of *Acc*I that resulted in incomplete digestion.

Search for Pseudogene Transcripts in Libraries

Three human cDNA libraries, normal skin fibroblasts, transformed myeloma cells, and teratocarcinoma cells, were also tested for pseudogenomic cDNAs, again with PsG up and PsG down as the primers. When the ~370-bp PCR fragment obtained from each sample was cleaved with *Acc*I and analyzed electrophoretically, the DNA from skin fibroblasts and myeloma cells produced only the three bands expected from PTMA cDNA. However, the teratocarcinoma library, after exhaustive digestion, invariably resulted in an *Acc*I fragment of ~320 bp, a size diagnostic for one of four pseudogenes (PTMAP2–5). Since this fragment represented 1–5% of the total amplified DNA, the fragment was purified electrophoretically and cloned using the *Kpn*I sites included in the primers. Upon examination of 19 cloned fragments, the inserts appeared to arise from fragments of PTMA cDNA that had apparently resisted cleavage at the downstream *Acc*I site; no inserts were derived from the pseudogenes. Therefore, the search for pseudogene transcripts failed. If any of the pseudogenes is expressed, the level of the mRNA is well below detectable levels in the materials examined, or the proper tissue has not yet been screened.

Concluding Remarks

The prothymosin α gene family consists of one intron-containing functional gene and five processed pseudogenes that, at the level of DNA, appear to be nearly perfect replicas of their parental counterpart. One of the latter encodes a protein virtually identical to authentic prothymosin α. In addition, four of the processed genes are adjacent to TATA boxes that meet the consensus requirements for sequence and spacing with respect to the transcriptional start site. The observation that processed pseudogenes frequently lie downstream of TATA-like elements is not surprising; they are usually inserted into A/T-rich loci. Since the prothymosin α pseudogenes genes are not transcriptionally active, the sequences in which their TATA boxes are embedded must be implicated in maintaining their silence. Therefore, the G/C-rich context in which active TATA boxes are located must be crucial to basal promoter activity.

The propagation of processed pseudogenes in vertebrate genomes requires that they originate from transcripts expressed in germ-line tissue (Vanin, 1985; Wilde, 1986). Consequently, it is likely that most, if not all, of the parental transcripts encode proteins involved in housekeeping activities. Since processed pseudogene formation is undoubtedly an ongoing process, one might ask how much additional DNA a genome can accept, and if tolerated, toward what end? In the prothymosin α gene family, 5 kb of DNA is devoted to expression, and an even greater amount (6 kb) to pseudogenes. Is this material evolutionarily neutral? Does it eventually serve to increase genetic diversity? Is there a benefit to preserving exons contiguously? Regardless of the answers to these questions, it is clear that nearly identical sequences scattered about the genome might increase the risk of nonreciprocal recombination. It is difficult to imagine retaining such potentially hazardous DNA if no advantage accrues. We speculate that processed pseudogenes have a function, even if that function remains obscure.

REFERENCES

Barberis, A., Superti-Furga, G., and Busslinger, M. (1987). Mutually exclusive interaction of the CCAAT-binding factor and of a displacement protein with overlapping sequences of a histone gene promoter. *Cell* **50:** 347–359.

Breathnach, R., and Chambon, P. (1981). Organization and expression of eukaryotic split genes coding for proteins. *Annu. Rev. Biochem.* **50:** 349–383.

Dang, C. V., and Lee, W. M. F. (1988). Identification of the human c-*myc* protein nuclear translocation signal. *Mol. Cell. Biol.* **8:** 4048–4054.

Dudow, K. P., and Perry, R. P. (1984). The gene family encoding the mouse ribosomal protein L32 contains a uniquely expressed intron-containing gene and an unmutated processed gene. *Cell* **37:** 457–468.

Eilers, M., Schirm, S., and Bishop, J. M. (1991). The *myc* protein activates the transcription of the α-prothymosin gene. *EMBO J.* **10:** 133–141.

Eschenfeldt, W. H., and Berger, S. L. (1986). The human prothymosin α gene is polymorphic and induced upon growth stimulation: Evidence using a cloned cDNA. *Proc. Natl. Acad. Sci. USA* **83:** 9403–9407.

Eschenfeldt, W. H., Manrow, R. E., Krug, M. S., and Berger, S. L. (1989). Isolation and partial sequencing of the human prothymosin α gene family: Evidence against export of the gene products. *J. Biol. Chem.* **264:** 7546–7555.

Gomez-Marquez, J., Segade, F., Dosil, M., Pichel, J. G., Bustelo, X. R., and Freire, M. (1989). The expression of prothymosin α gene in T lymphocytes and leukemic lymphoid cells is tied to lymphocyte proliferation. *J. Biol. Chem.* **264:** 8451–8454.

Goodall, G. J., Dominguez, F., and Horecker, B. L. (1986). Molecular cloning of cDNA for human prothymosin α. *Proc. Natl. Acad. Sci. USA* **83:** 8926–8928.

Guiochon-Mantel, A., Loosfelt, H., Lescop, P., Sar, S., Atger, M., Perrot-Applanat, M., and Milgrom, E. (1989). Mechanisms of nuclear localization of the progesterone receptor: Evidence for interaction between monomers. *Cell* **57:** 1147–1154.23.

Haritos, A. A., Goodall, G. J., and Horecker, B. L. (1985). Prothymosin α and α_1-like peptides. *In* "Methods in Enzymology" (G. Di Sabato, J. J. Langone, and H. Van Vunakis, Eds.), Vol. 116, pp. 255–265, Academic Press, New York.

Kleinschmidt, J. A., and Seiter, A. (1988). Identification of domains involved in nuclear uptake and histone binding of protein N1 of *Xenopus laevis*. *EMBO J.* **7:** 1605–1614.

Kozak, M. (1986). Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* **44:** 283–292.

Kozak, M. (1987). At least six nucleotides preceding the AUG initiator codon enhance translation in mammalian cells. *J. Mol. Biol.* **196:** 947–950.

Lewin, B. (1985). "Genes II," pp. 196–201, Wiley, New York.

Low, T. L. K., and Goldstein, A. L. (1985). Thymic hormones: An overview. *In* "Methods in Enzymology" (G. Di Sabato, J. J. Langone, and H. Van Vunakis, Eds.), Vol. 116, pp. 213–215, Academic Press, New York.

Manrow, R. E., Sburlati, A. R., Hanover, J. A., and Berger, S. L. (1991). Nuclear targeting of prothymosin α. *J. Biol. Chem.* **266:** 3916–3924.

Ogden, R. C., and Adams, D. A. (1987). Electrophoresis in agarose and acrylamide gels. *In* "Methods in Enzymology" (S. L. Berger and A. R. Kimmel, Eds.), Vol. 152, pp. 61–87, Academic Press, San Diego.

Richardson, W. D., Roberts, B. L., and Smith, A. E. (1986). Nuclear location signals in polyoma virus large-T. *Cell* **44:** 77–85.

Sanger, F. (1981). Determination of nucleotide sequences in DNA. *Science* **214:** 1205–1210.

Sasavage, N. L., Smith, M., Gillam, S., Woychik, R. P., and Rottman, F. M. (1982). Variation in the polyadenylation site of bovine prolactin mRNA. *Proc. Natl. Acad. Sci. USA* **79:** 223–227.

Sburlati, A. R., Manrow, R. E., and Berger, S. L. (1991). Prothymosin α antisense oligomers inhibit myeloma cell division. *Proc. Natl. Acad. Sci. USA* **88:** 253–257.

Schmid, C. W., and Jelinek, W. R. (1982). The *Alu* family of dispersed repetetive sequences. *Science* **216:** 1065–1070.

Shelness, G. S., and Williams, D. L. (1984). Apolipoprotein II messenger RNA: Transcriptional and splicing heterogeneity yields six 5'-untranslated leader sequences. *J. Biol. Chem.* **259:** 9929–9935.

Silver, J., and Keerikatte, V. (1989). Novel use of the polymerase chain reaction to amplify cellular DNA adjacent to an integrated provirus. *J. Virol.* **63:** 1924–1928.

Siomi, H., Shida, H., Nam, S. K., Nosaka, T., Maki, M., and Hatanaka, M. (1988). Sequence requirements for nucleolar localization of human T cell leukemia virus Type I pX protein, which regulates viral RNA processing. *Cell* **55:** 197–209.

Skowronski, J., Fanning, T. G., and Singer, M. F. (1988). Unit-length line-1 transcripts in human teratocarcinoma cells. *Mol. Cell. Biol.* **8:** 1385–1397.

Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98:** 503–517.

Vanin, E. F. (1985). Processed pseudogenes: Characteristics and evolution. *Annu. Rev. Genet.* **19:** 253–272.

Wilde, D. (1986). Pseudogenes. *CRC Rev. Biochem.* **19:** 323–352.