MOLECULAR AND CELLULAR BIOLOGY, July 1985, p. 1694–1706
0270-7306/85/071694-13$02.00/0

Vol. 5, No. 7

Characterization of the Functional Gene and Several Processed Pseudogenes in the Human Triosephosphate Isomerase Gene Family

JUDITH R. BROWN, IRA O. DAAR, JAMES R. KRUG, AND LYNNE E. MAQUAT*

Department of Human Genetics, Roswell Park Memorial Institute, New York State Department of Health, Buffalo, New York 14263

Received 31 January 1985/Accepted 2 April 1985

The functional gene and three intronless pseudogenes for human triosephosphate isomerase were isolated from a recombinant DNA library and characterized in detail. The functional gene spans 3.5 kilobase pairs and is split into seven exons. Its promoter contains putative TATA and CCAAT boxes and is extremely rich in G and C residues (76%). The pseudogenes share a high degree of homology with the functional gene but contain mutations that preclude the synthesis of an active triosephosphate isomerase enzyme. Sequence divergence calculations indicate that these pseudogenes arose approximately 18 million years ago. We present evidence that there is a single functional gene in the human triosephosphate isomerase gene family.

Triosephosphate isomerase (TPI, EC 5.3.1.1) catalyzes the interconversion of dihydroxyacetone phosphate and glyceraldehyde-3-phosphate in the glycolytic and gluconeogenic pathways (40). The enzyme consists of two identical polypeptide chains which are 248 amino acids in humans (24, 27). Multiple electrophoretic and chromatographic forms have been reported for TPI in every human tissue that has been examined (21, 42, 51). However, gene mapping studies with somatic cell hybrids (18, 23), studies of TPI posttranslational deamidation (15, 36), and characterizations of naturally occurring variant isozyme patterns (12, 42) indicate that the multiple forms arise from posttranslational modifications of a protein that is encoded by a single gene.

TPI is required for cell growth and maintenance. Accordingly, the TPI gene is expressed in all cell types and thus belongs to the so-called housekeeping gene category. Most previous studies of mammalian gene regulation have focused on facultative genes whose expression is limited to a particular cell type and developmental stage. In contrast, little is known about housekeeping gene regulation. We have undertaken a characterization of the human TPI gene to improve our understanding of the structural features that distinguish housekeeping genes from developmentally controlled genes. Studies of this gene should define DNA sequences that promote housekeeping gene recognition by the transcriptional machinery of all cells.

Another interest in the TPI gene involves the various mutations that cause hereditary TPI deficiencies in humans. Homozygous-deficient individuals usually have 3 to 20% of normal TPI activity and exhibit chronic nonspherocytic hemolytic anemia, retarded growth, increased susceptibility to bacterial infections, and pronounced neurological and muscular disorders (54). The mutations that cause TPI deficiency appear to be heterogeneous since both TPI enzyme activity and steady-state TPI mRNA levels can vary among patients (27). For those patients tested, residual enzyme activity is heat labile, suggesting that at least one allele encodes a structurally altered protein (27, 49, 56). Although the incidence of homozygous TPI deficiency is fairly rare, a significant percentage of the human population (e.g., approximately 5% of Blacks in the United States) carries one normal and one null TPI allele as defined by 50% enzyme activity and 50% immunological cross-reacting material (13, 34–36, 47). These individuals are phenotypically normal. Characterization of mutant alleles that alter steady-state TPI mRNA levels as well as characterization of null alleles that do not produce detectable protein should help to identify nucleotides that function in the control of TPI gene transcription and RNA processing.

Nine human TPI cDNA sequences were isolated from an adult liver library (27). All cDNAs appear to be derived from a single mRNA species, suggesting derivation from transcripts of a single gene. DNA sequencing defined the entire 744-nucleotide coding region, from which the TPI amino acid sequence was deduced, and the entire 448-nucleotide 3′ untranslated region. Hybridization of TPI cDNA to restriction enzyme digests of human DNA demonstrated that the human genome contains multiple copies of TPI gene sequences (27).

We report here the isolation and characterization of the functional gene and three pseudogenes for human TPI. The functional gene spans 3.5 kilobase pairs (kbp) and is split into seven exons. Exon-intron boundaries and the putative promoter region were defined by nucleotide sequence analysis. The 5′ end of TPI mRNA was localized by primer extension. The three pseudogenes lack introns and evolved independently from processed transcripts of the functional gene approximately 18 million years ago.

MATERIALS AND METHODS

Cell cultures. Monolayer cultures of L153 and WI38 diploid fibroblast cell lines and the HeLa cell line were grown in Dulbecco modified Eagle medium supplemented with 10% fetal calf serum. The cells were chilled to 4°C when 80% confluent and then harvested by scraping. Daudi Burkitt lymphoma B cells, Jurkat-FHCRC acute lymphocytic leukemia T cells, and Frawley Epstein-Barr virus-infected B cells were grown in suspension culture in RPMI 1640 supplemented with 10% fetal calf serum.

Genomic DNA library screening. Approximately 600,000 plaques of a human genomic DNA library in λ Charon 4A (26) were screened (6) with a nick-translated, *Pst*I-excised insert from the human cDNA clone pHTPI-5a (27). Nitrocellulose-bound DNA was prehybridized at 37°C in a solution containing 5× SSC (1× SSC is 0.15 M NaCl plus 0.015 sodium citrate), 50% formamide, 1× Denhardt solution, 20

* Corresponding author.

mM sodium phosphate (pH 7.5), and 50 μg of single-stranded calf thymus DNA per ml. cDNA was denatured and added to 1.5×10^6 cpm/ml. After the hybridization, filters were washed at 50°C in 0.1× SSC–0.1% sodium dodecyl sulfate and exposed to X-ray film.

Restriction enzyme mapping the TPI genomic clones in bacteriophage DNA. Physical maps of each genomic clone in λ Charon 4A were initially generated by localizing all *Eco*RI and *Pst*I sites relative to the right arm of the recombinant molecules. DNA was digested to various degrees of completion with the appropriate enzyme, electrophoresed in a 0.3% agarose gel, transferred to nitrocellulose (52), and prehybridized at 37°C in a solution containing 5× SSC, 50% formamide, 10% dextran sulfate, 1× Denhardt solution, 20 mM sodium phosphate (pH 7.5), 50 μg of single-stranded calf thymus DNA per ml, and 0.1% sodium dodecyl sulfate. A 1,980-base-pair (bp) *Bgl*II-*Bam*HI fragment derived from the right arm of λ Charon 4A DNA (bp 41,890 to 43,870) was labeled with ^{32}P by nick translation and added to 1.5×10^6 cpm/ml. After hybridization, the nitrocellulose was washed at 50°C in 0.1× SSC–0.1% sodium dodecyl sulfate and exposed to X-ray film. For more detailed analysis, genomic fragments generated by complete digestion with one or more restriction enzymes were subcloned into pBR322, pUC13, or M13 DNA.

Blot hybridization of genomic DNAs. High-molecular-weight chromosomal DNA was prepared from cultured cells by phenol extraction and CsCl gradient centrifugation (45). DNA was digested with restriction enzymes, electrophoresed in a 0.8% agarose gel, transferred to nitrocellulose, and hybridized to ^{32}P-labeled TPI cDNA as described above for restriction enzyme mapping. To identify restriction fragments containing intron sequences, a 245-bp *Hin*fI-*Pst*I fragment derived from intron 4 of the hTPI-8B gene was subcloned into the *Hin*cII and *Pst*I sites of M13mp8 DNA. Before ligation, the *Hin*fI site in intron 4 was filled in with Klenow fragment and deoxynucleoside triphosphates. The single-stranded subclone was labeled with ^{32}P by hybridization to and extension of the M13 universal primer. Double-stranded [^{32}P]DNA was digested with *Pst*I and *Eco*RI, and the TPI insert was purified. Hybridization to filter-bound DNA was as described above for restriction enzyme mapping.

DNA sequencing. DNA restriction fragments were sequenced by either the chemical method of Maxam and Gilbert (28) or the dideoxy termination method of Sanger et al. (46).

RNA blotting. Total RNA was isolated from cultured cells (45). Polyadenylated [poly(A)$^+$] RNA was purified by chromatography on oligodeoxythymidylic acid-cellulose (2), electrophoresed in a 1.2% agarose–2.2 M formaldehyde slab gel, transferred to nitrocellulose (53), and hybridized to a ^{32}P-labeled subclone of TPI cDNA (27).

Primer extension analysis. A 284-bp *Hin*fI fragment that includes nucleotides −221 to +63 (where +1 is the first nucleotide of the TPI translation initiation codon) was isolated from a subclone of the hTPI-8B gene, 5′ labeled with [γ-^{32}P]ATP and polynucleotide kinase, digested with *Hae*II, and electrophoresed in a 20% polyacrylamide strand-separating gel (25). The ^{32}P-labeled, mRNA-complementary strand that includes nucleotides +4 to +63 was purified and used as the primer. Poly(A)$^+$ RNA (1 μg) isolated from Frawley Epstein-Barr virus-infected B cells, and 25 ng of ^{32}P-labeled primer was suspended in 80% phosphate-buffered formamide–7.5 mM Tris (pH 7.5)–0.45 M NaCl–0.1 mM EDTA and denatured at 70°C. After 5 min, the temperature was gradually lowered to 37°C to maximize hybridization. The hybrids were then ethanol precipitated, and the primers were extended with reverse transcriptase at 42°C in 50 mM Tris (pH 8.1)–2 mM dithiothreitol–5 mM $MgCl_2$–10 mM KCl–0.6 mM deoxynucleoside triphosphates. After 90 min, NaOH was added to 0.2 M, and the incubation was continued at 42°C for an additional 2 h. The reactions were neutralized by the addition of Tris (pH 7.0) to 0.7 M, ethanol precipitated, and denatured in 100% formamide. Extended primers were sized by electrophoresis in an 8% polyacrylamide–7 M urea gel and detected by autoradiography.

Gene divergence calculations. Human pseudogene or yeast functional gene sequences (1) were aligned with the human functional gene sequence to maximize similarities. Nucleotide substitutions per coding region site (percent divergence divided by 100) were determined by the method of Perler et al. (41). This method introduces corrections for multiple events at a single site. All pseudogene codons harboring deletions or insertions relative to functional gene codons were excluded from the calculations. Times of pseudogene divergence from the functional gene were determined by using the average silent site substitution rates calculated for globin genes (14, 22, 41). When comparing pseudogene and functional gene untranslated regions, a nucleotide in common to the two regions was scored as +1, a substitution, deletion, or insertion was scored as −1, and the denominator (total number of nucleotides evaluated) was taken as the average number of nucleotides in the two sequences under comparison. Corrections for multiple events at a single site within the untranslated regions were made by using the formulation of Jukes and Cantor (19).

RESULTS

Isolation and characterization of human TPI genomic clones. Several human TPI cDNA sequences have been isolated from an adult liver cDNA library (27). All appear to be derived from a single mRNA species. The longest cDNA, pHTPI-5a, contains the last two nucleotides of the translation initiation codon, the entire 744-nucleotide coding region of the mature polypeptide, and the entire 448-nucleotide 3′ untranslated region. The purified cDNA insert of pHTPI-5a was ^{32}P labeled by nick translation and used to isolate TPI sequences from a twice-amplified human genomic library in λ Charon 4A (26). Of the 600,000 plaques that were screened (6), 4 contained different TPI sequences as determined by restriction enzyme analysis of phage DNA. cDNA-homologous regions were mapped to specific restriction enzyme fragments of each recombinant by blot hybridization. Exon sequences were more precisely localized by DNA sequencing.

Restriction enzyme maps of the four genomic clones have few similarities (Fig. 1). However, all clones share an *Nco*I site that lies close to the 5′ end of the TPI coding region and, as determined by DNA sequencing (see below), includes the ATG translation initiation codon. cDNA-homologous sequences in λhTPI-8B are dispersed over several kbp. In contrast, cDNA-homologous sequences in λhTPI-5A, λhTPI-19A, and λhTPI-13C are clustered within an approximately 1-kbp region.

Structure of the functional TPI gene. The restriction map of λhTPI-8B DNA suggested that cDNA-homologous sequences in the hTPI-8B gene were interrupted by introns. Nucleotide sequence analysis confirmed that this gene consists of seven exons and six introns that span 3.5 kbp. The exons contain the entire TPI protein-coding region and predict an amino acid sequence that is identical to that

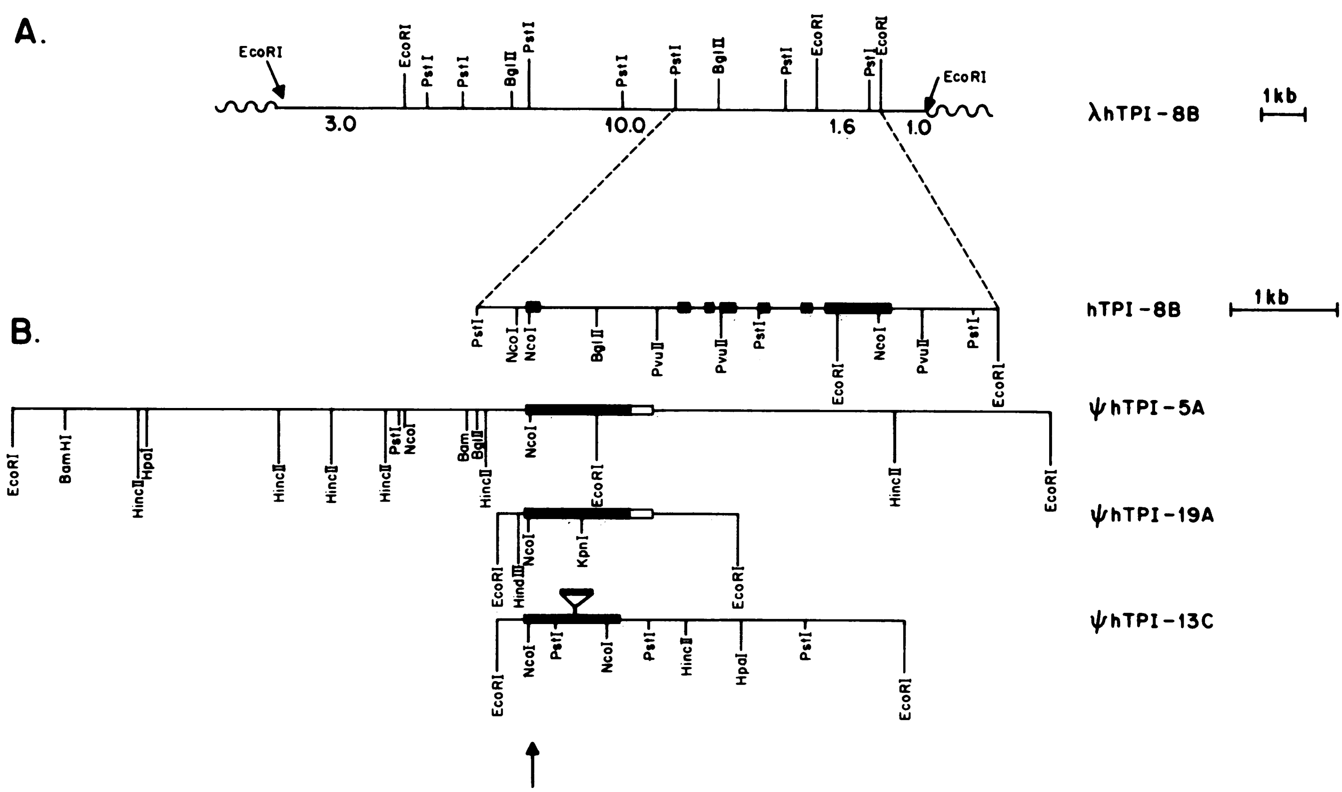


FIG. 1. Physical maps of TPI genomic sequences. (A) Organization of human DNA in λhTPI-8B. The linkage maps of *Eco*RI and *Pst*I fragments were determined by partial restriction enzyme digests. The horizontal line represents human DNA, and the wavy line represents λ DNA. *Eco*RI sites at the insert-vector junctions were created in the cloning procedure and do not reflect *Eco*RI sites in human DNA (20). Numbers below the horizontal line indicate *Eco*RI fragment sizes in kbp. (B) Localization of TPI gene sequences in recombinant bacteriophage DNA. Only *Eco*RI fragments or, as in the case of hTPI-8B, a part of an *Eco*RI fragment, that contain TPI gene sequences are shown. All genes are oriented 5′ to 3′ and are aligned by using the conserved *Nco*I site that includes the ATG translation initiation codon. This site is indicated by the arrow below the figure. Solid boxes designate exon sequences. Open boxes in the ΨhTPI-5A and ΨhTPI-19A genes represent exon regions that presumably exist yet have not been rigorously analyzed by DNA sequencing. The Y in the ΨhTPI-13C gene designates the 317-bp sequence that most likely was deleted by a homologous recombination event. Note that the *Eco*RI site within the ΨhTPI-5A gene corresponds to the *Eco*RI site in the last exon of the hTPI-8B gene and that the second *Nco*I site within the ΨhTPI-13C gene corresponds to the *Nco*I site in the last exon of the hTPI-8B gene.

previously determined from pHTPI-5a cDNA (Fig. 2). Although there are two nucleotide differences between the exon regions of the hTPI-8B gene and pHTPI-5a cDNA, these differences are silent mutations. One mutation is in nucleotide 3 of codon 94, and the other is in nucleotide 3 of codon 162. Silent site substitutions between hTPI-8B and pHTPI-5a DNA are not unexpected since these DNAs were derived from different individuals. The 3′ untranslated regions of hTPI-8B and pHTPI-5a are completely homologous.

The hTPI-8B gene sequences at exon-intron boundaries (Table 1) are in reasonable agreement with the consensus sequences established for other functional genes that are transcribed by RNA polymerase II (38). Two exon-intron junctions lie between codons, three junctions interrupt a codon after the first nucleotide, and one junction interrupts a codon after the second nucleotide. All introns are within the protein-coding portion of the gene. Exons range in size from 74 to 564 bp, the largest of which includes the 448-bp 3′ untranslated region.

TPI mRNA structure. A low-abundance, 1,250-nucleotide RNA species was detected by blot hybridization of total RNA isolated from human fibroblasts, HeLa cells, Daudi Burkitt lymphoma B cells, Jurkat-FHCRC acute lymphocytic leukemia T cells, and Frawley Epstein-Barr virus-infected B cells (Fig. 3; 27). Variations in the relative abundance of TPI RNA between cell lines may reflect different cellular growth rates. Given a 744-nucleotide coding region and a 448-nucleotide 3′ untranslated region (as defined by cDNA and genomic DNA sequencing) and assuming a 50-nucleotide poly(A) tail, the 5′ untranslated region of TPI mRNA must be no larger than 50 nucleotides.

The 5′ terminus of TPI mRNA was localized by primer extension analysis. A single-stranded, 5′-[^{32}P]DNA primer complementary to nucleotides +4 to +63 (Fig. 4; +1 is the first nucleotide of the translation initiation codon) was hybridized to poly(A)$^+$ RNA and extended with deoxynucleoside triphosphates by reverse transcriptase. Extension products were analyzed by denaturing gel electrophoresis and autoradiography. Products corresponding to termini at guanosine nucleotides −32 and −34 (which define a 5′ untranslated region of 32 and 34 bp, respectively) were detected with Frawley Epstein-Barr virus-infected B-cell RNA (Fig. 4) and Daudi Burkitt lymphoma B-cell RNA (data not shown). Since the mRNA cap structure may inhibit reverse transcription of the last few nucleotides, these 5′-end determinations are only close approximations, and the smaller product may be the result of premature transcription termination. Alternatively, this product could reflect a second 5′ end for TPI mRNA. The fact that homology of the functional gene with at least two of the processed TPI pseudogenes stops abruptly upstream from position −34 (see below) strongly suggests that the longer primer extension product reflects the actual 5′ end of TPI mRNA. Although an adenine nucleotide is the most common nucleotide at an mRNA start site, transcription initiation at guanine nucleotides has been reported for other genes (7). Adenine nucleotides in closest proximity to position −34 reside at positions −39 and −29.

Features of the TPI gene promoter region. Sequences upstream from the proposed mRNA start site of the hTPI-8B gene contain putative TATA and CCAAT boxes at positions that are consistent with the usual spacings of a eucaryotic promoter (11). The TATA box (TATATA) is located 26 to 21 bp and the CCAAT box (CCAT) is located 73 to 70 bp 5′ to the proposed site of transcription initiation. The most outstanding feature of sequences upstream from the TPI gene is a high G-C content. Positions −35 to −140 are 76% G and C nucleotides, with a preponderance (58%) of G nucleotides. The 5′ untranslated region is also notably G-C rich (73%).

Structure of three TPI pseudogenes. To investigate the structure of other members of the TPI gene family, cDNA-homologous sequences were determined in entirety for λhTPI-13C and in part for λhTPI-19A and λhTPI-5A (Fig. 2). All three genes harbor multiple amino acid substitutions as well as insertions and deletions that preclude the synthesis of a functional TPI polypeptide. Deletions in λhTPI-19A and λhTPI-5A shift the translational reading frame and generate premature termination codons. All three genes are intronless and thus bear a structural hallmark of so-called processed pseudogenes. The ΨhTPI-13C gene has a tract of eight adenine nucleotides that lies downstream from the AATAAA sequence in the 3′ untranslated region. In addition, this gene is flanked by a short direct repeat, TAAATTT. These features suggest that the ΨhTPI-13C gene was generated by integration of a reverse-transcribed copy of TPI mRNA into germ line DNA via a transposition-like mechanism (17, 39, 55). Extrapolating from the structure of other processed pseudogenes, the position of the 5′-flanking repeat often delineates the 5′ end of mRNA sequences. The 5′-most repeat in the ΨhTPI-13C gene is immediately upstream from the site proposed for functional gene transcription initiation. Therefore, in agreement with mRNA primer extension analysis, the ΨhTPI-13C gene structure indicates that TPI mRNA consists of a 34-nucleotide 5′ untranslated region. The 3′-most repeat in ΨhTPI-13C is four nucleotides 3′ to the 8-bp stretch of adenine nucleotides that was probably derived by reverse transcription of a portion of the TPI mRNA poly(A) tail.

Although nucleotide sequences at the 3′ end of the ΨhTPI-19A and ΨhTPI-5A genes have not been determined, the fact that at least ΨhTPI-19A and possibly ΨhTPI-5A diverge from the functional gene at the same 5′ position (−34) as does ΨhTPI-13C indicates that they too arose from germ line integration of a TPI cDNA. In accordance with primer extension data, we assume that the homology of ΨhTPI-5A with hTPI-8B at positions −36 and −35 is fortuitous rather than reflecting heterogeneity in the 5′ terminus of TPI mRNA. By analogy to the ΨhTPI-13C gene, ΨhTPI-19A and ΨhTPI-5A sequences immediately upstream from bp −34 should include one of the flanking direct repeats that are remnants of the cDNA integration event. Since sequences flanking the ΨhTPI-13C, ΨhTPI-19A, and ΨhTPI-5A genes bear no relationship to one another, each gene was integrated into a different chromosome locus (see below).

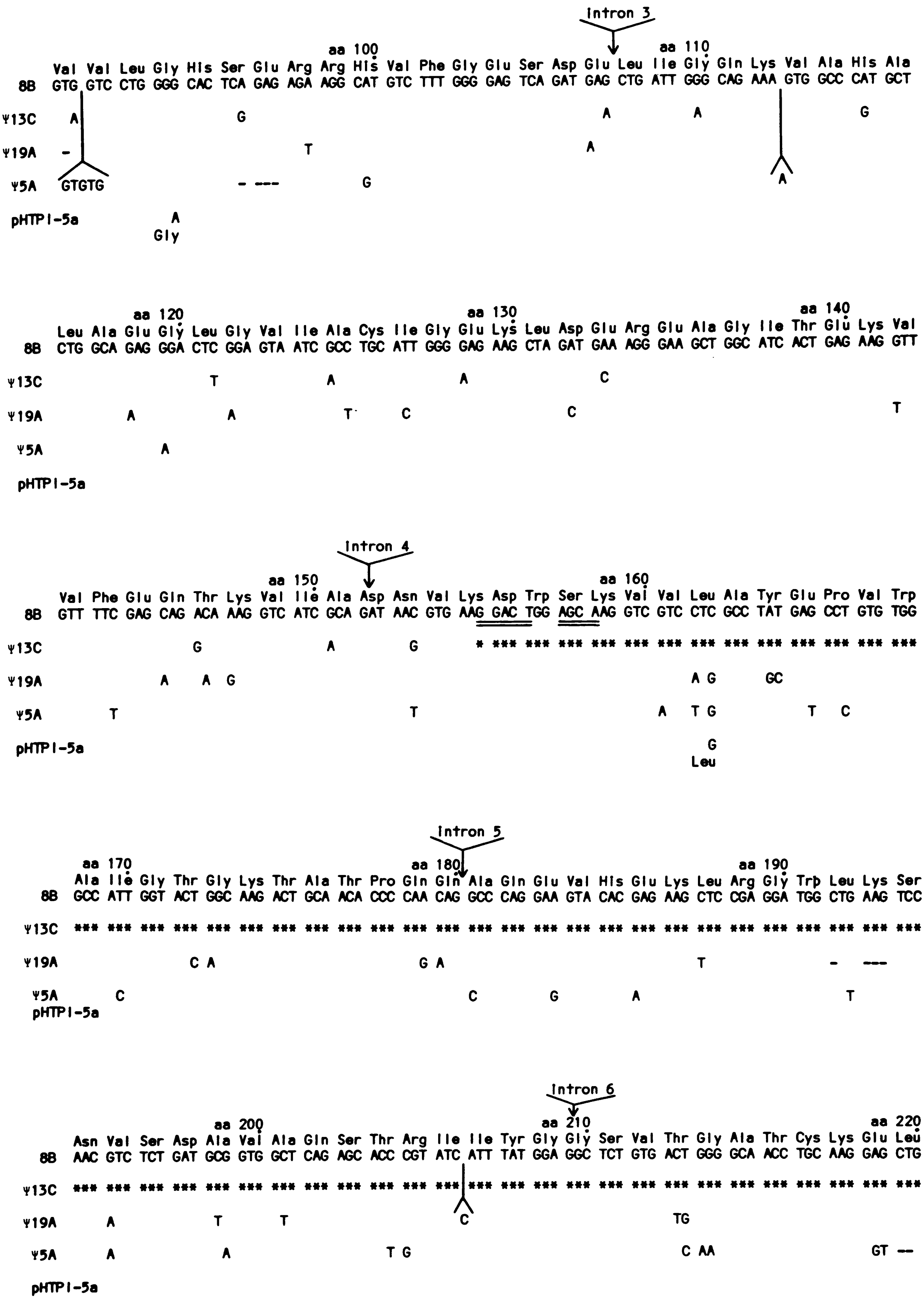
The ΨhTPI-13C gene has a 317-bp deletion (designated by asterisks in Fig. 2) that extends from codon 155 to 40 bp into the 3′ untranslated region. An examination of functional gene sequences near the deletion endpoints reveals a direct repeat of GGACT$(N)_1$ or $_2$AGCA. The equivalent of one of these repeats, GGACTAAGCA, remains in ΨhTPI-13C DNA, suggesting that the deletion arose by homologous recombination between the repeats either in human cells before library construction or during λΨhTPI-13C propagation in *Escherichia coli*. Deletion endpoints are indeterminable because the exact site of crossover is unknown. For simplicity, the deletion is shown to include all of the first repeat and none of the second (Fig. 2). This is consistent with the GGACTAAGCA sequence that remains in the cloned pseudogene.

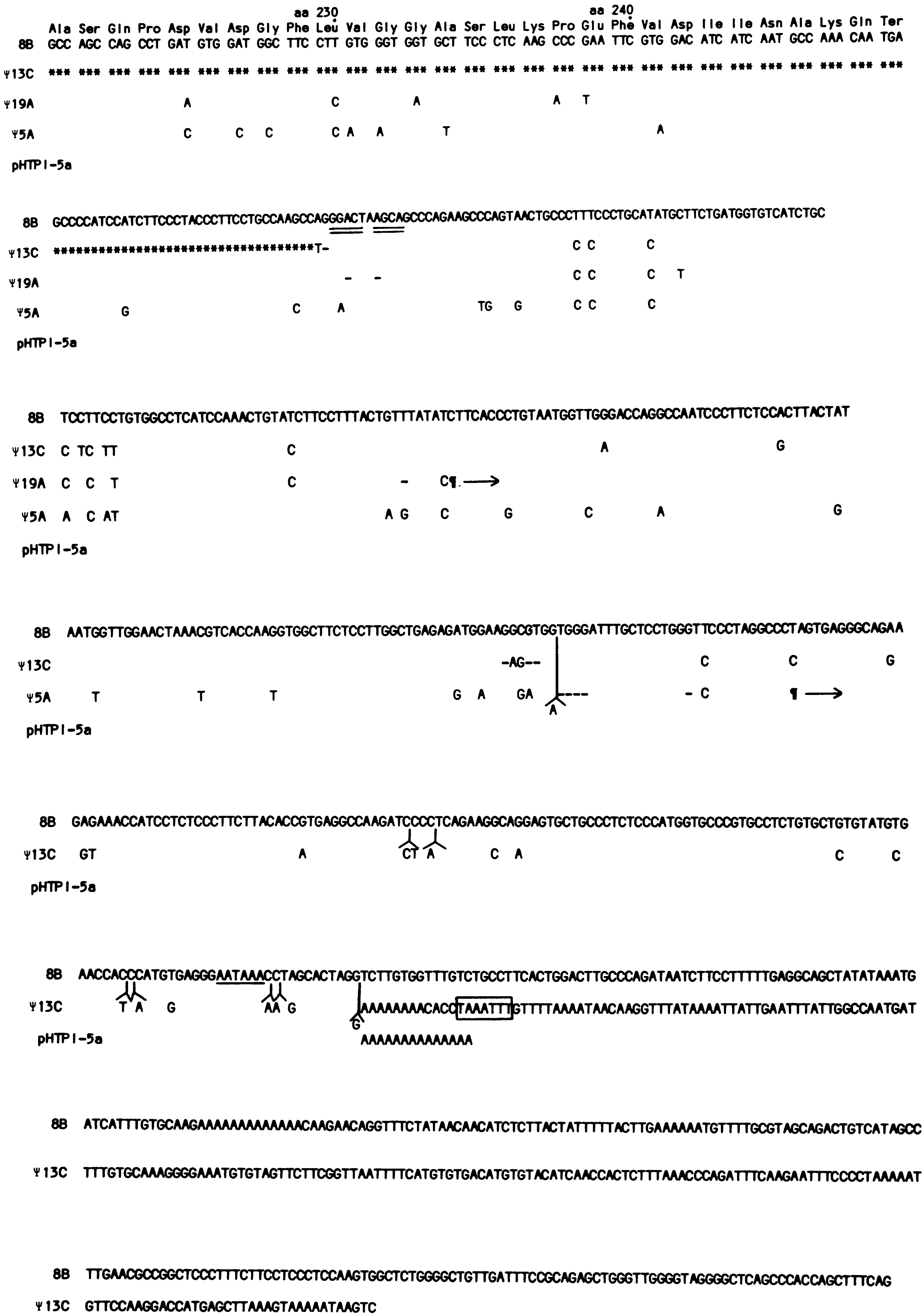
Genomic organization of the human TPI gene family. TPI sequences in DNA from two human cell lines were examined to estimate the number of genes comprising the human TPI gene family and to associate a particular genomic DNA restriction fragment with a particular TPI gene. Genomic DNA and TPI recombinant bacteriophage DNA were incubated with *Pst*I or *Eco*RI, electrophoresed in an agarose gel, transferred to nitrocellulose, and hybridized with a nick-translated cDNA insert from pHTPI-5a. Of the approximately nine *Pst*I genomic DNA fragments that were detected, three could be correlated with one of the characterized TPI genes on the criterion that they comigrate with *Pst*I fragments in recombinant phage DNA (Fig. 5A). In L153 and HeLa cell DNA, the entire functional hTPI-8B gene is contained within 2.6- and 2.0-kbp *Pst*I fragments. Only a part of the genomic *Pst*I fragment that harbors the ΨhTPI-5A gene is present in λΨhTPI-5A. Similarly, only a part of the genomic *Pst*I fragment that harbors the ΨhTPI-19A gene is present in λΨhTPI-19A. Therefore, neither of these genes can be associated with a *Pst*I fragment of cell line DNA by this method. Human DNA appears to contain a *Pst*I fragment of similar size to the 0.9-kbp *Pst*I fragment that spans the 317-bp deletion in the ΨhTPI-13C gene. This suggests but does not prove that the deletion in λΨhTPI-13C DNA arose in human cells as opposed to *E. coli*. HeLa cell DNA (and L153 fibroblast DNA [data not shown]) differs from W138 cell DNA by two additional *Pst*I fragments of approximately 1.0 and 1.7 kbp.

*Eco*RI restriction fragments of genomic and TPI recombinant bacteriophage DNA were also analyzed. Of the approximately nine *Eco*RI genomic DNA fragments that hybridize to TPI cDNA, five comigrate with *Eco*RI fragments in recombinant phage DNA (Fig. 5B). The entire functional hTPI-8B is present within 10.0- and 1.6-kbp *Eco*RI fragments, the entire ΨhTPI-5A gene is present within 5.7- and 4.3-kbp *Eco*RI fragments, and the entire ΨhTPI-19A gene is contained within a 2.3-kbp *Eco*RI fragment. Neither the deletion-bearing 3.9-kbp *Eco*RI fragment of λΨhTPI-13C nor an *Eco*RI fragment approximately 300 bp larger is detected in human DNA. This is expected because the 3.9-kbp fragment maps to a human-λ DNA junction in λΨhTPI-13C, and therefore, one of the *Eco*RI ends of this fragment was derived from an *Eco*RI linker used in the cloning process (26).

We estimate from the number of genomic restriction fragments that hybridize to TPI cDNA that there are two or three TPI genes in addition to those characterized. To determine whether any of these genes contain intron sequences, 245 nucleotides of intron 4 were subcloned and hybridized to *Pst*I-digested genomic DNA. Intron 4 sequences were detected only in the 2.6-kbp *Pst*I fragment that was shown to contain the hTPI-8B gene (Fig. 6). We conclude that the hTPI-8B gene is the only intron-containing

```
       -360      -350      -340      -330      -320      -310      -300      -290      -280      -270
8B    ACCAGCTAGTTCCGCTTGAAAACCACTTCTGGCCCCGTGGGGGACTCAAGTCGCCAAGCGAGGGTTCCCCTGAGCGCCGGAGCTCACAGGTCTCGCCTTGTCCCG

     -260      -250      -240      -230      -220      -210      -200      -190      -180      -170      -160
8B    AAAGCCCCGCAATCGAGGCGGAGGCGACCGAGCCCCCGACTCTCCTAGAACGTTGCCACAAGAAGGGGGAACGTCGGAACAGTGCATCATCGGGCGGCGGCCGGG
Ψ13C <-¶CCAGGTTCAATCAGAATGGTAAATTGTACAAATATATGCACATATATATACCATATAAAATTCTATTTCAAGGTTTTTAATTACTTACCACTTTGCCCAGTTT
Ψ5A            <--- ¶ACAAAAGAAAACAGAAATATTGGAAACTGTACTATGGAGAAATTAGGGACAAAAGGTAACAGTATATTGATATTAACATTGCTGCTAGT

       -150      -140      -130      -120      -110      -100       -90       -80       -70       -60
8B    GCGGCGGCAGGAGGGCGGGCGGGGGGCAGGGCTCCGGGGGACTGGGCGGGCCATGGCGGAGGACGGCGAGGAGGCGGAGTTCCACTTCGCGGCGCTCTATATAA
Ψ13C  TCTCTCATAGGCTTTATGCCGTAGTCCTAATCCCAACAATCCCACACCAACAAGCAGCCACAGAACTAAACAGAAAAAAGAAATAATTGGGTTTTTTTAAAATA
Ψ19A                                   <--- ¶TCAGGTTCTTACATTTTTTTTAACATGATACCTTTTAGAATATCAAAAACCTATGATCCCC
Ψ5A   CCTTTGCACTAGTAAATAACTGCTATTTGATAAATGATCACAATGTGTAAAACACTGTAGTTACAAGATCTCATTTAATCCGCCTAACAACCTTGCCAAGTATT

                        r>                                        aa 1                               aa 10
       -50       -40    : -30       -20       -10             Met Ala Pro Ser Arg Lys Phe Phe Val Gly Gly Asn Trp
8B    GTGGGCAGTGGCCGACTGCGCGCGACACTGACCTTCAGCGCCTCGGCTCGGCC ATG GCG CCC TCC AGG AAG TTC TTC GTT GGG GGA AAC TGG
Ψ13C  AATAAAATAAAA[TAAATTT] -T          T     A    T         A                       -    A     -C
                                                  CAG
Ψ19A  AACCTAGAAATCACAGAAG   -T          T     A    C            -                ---        G
                                                  CAG
Ψ5A   AATAAAACCCGTTTTAGGC   -T      A   T     A    C            ---  T               G    A G
                                                  CAG
pHTP1-5a

                                                                                               Intron 1
                          aa 20                                   aa 30                              |
     Lys Met Asn Gly Arg Lys Gln Ser Leu Gly Glu Leu Ile Gly Thr Leu Asn Ala Ala Lys Val Pro Ala Asp Thr Glu
8B   AAG ATG AAC GGG CGG AAG CAG AGT CTG GGG GAG CTC ATC GGC ACT CTG AAC GCG GCC AAG GTG CCG GCC GAC ACC GAG
Ψ13C           T   T       A       T               T A      —       T A          A       G T
Ψ19A             C A       A   C                 C         G                 -     A     T
Ψ5A                        A A T                           A            CT    T           A
pHTP1-5a

     aa 40                                aa 50                                       aa 60
     Val Val Cys Ala Pro Pro Thr Ala Tyr Ile Asp Phe Ala Arg Gln Lys Leu Asp Pro Lys Ile Ala Val Ala Ala Gln
8B   GTG GTT TGT GCT CCC CCT ACT GCC TAT ATC GAC TTC GCC CGG CAG AAG CTA GAT CCC AAG ATT GCT GTG GCT GCG CAG
Ψ13C           C T     A                 A A                                       C
Ψ19A           C A   T   C               T     —         G
Ψ5A    A             T G -   G   A   G   G                  C                         C  AC
pHTP1-5a

                                                      Intron 2
                                                         |
                    aa 70                                  aa 80                                  aa 90
     Asn Cys Tyr Lys Val Thr Asn Gly Ala Phe Thr Gly Glu Ile Ser Pro Gly Met Ile Lys Asp Cys Gly Ala Thr Trp
8B   AAC TGC TAC AAA GTG ACT AAT GGG GCT TTT ACT GGG GAG ATC AGC CCT GGC ATG ATC AAA GAC TGC GGA GCC ACG TGG
Ψ13C                             C                             A                 A   -
Ψ19A                 A     C     C     T                                         AA     AT C T
Ψ5A                              C                       T       G         TA    T
pHTP1-5a
```


Intron 3
aa 100
aa 110
Val Val Leu Gly His Ser Glu Arg Arg His Val Phe Gly Glu Ser Asp Glu Leu Ile Gly Gln Lys Val Ala His Ala
8B GTG GTC CTG GGG CAC TCA GAG AGA AGG CAT GTC TTT GGG GAG TCA GAT GAG CTG ATT GGG CAG AAA GTG GCC CAT GCT
Ψ13C A G A A G
Ψ19A - T A
Ψ5A GTGTG - --- G A
pHTPI-5a A
Gly
aa 120
aa 130
aa 140
Leu Ala Glu Gly Leu Gly Val Ile Ala Cys Ile Gly Glu Lys Leu Asp Glu Arg Glu Ala Gly Ile Thr Glu Lys Val
8B CTG GCA GAG GGA CTC GGA GTA ATC GCC TGC ATT GGG GAG AAG CTA GAT GAA AGG GAA GCT GGC ATC ACT GAG AAG GTT
Ψ13C T A A C
Ψ19A A A T C C T
Ψ5A A
pHTPI-5a
Intron 4
aa 150
aa 160
Val Phe Glu Gln Thr Lys Val Ile Ala Asp Asn Val Lys Asp Trp Ser Lys Val Val Leu Ala Tyr Glu Pro Val Trp
8B GTT TTC GAG CAG ACA AAG GTC ATC GCA GAT AAC GTG AAG GAC TGG AGC AAG GTC GTC CTC GCC TAT GAG CCT GTG TGG
Ψ13C G A G * *** *** *** *** *** *** *** *** *** *** *** *** *** ***
Ψ19A A A G A G GC
Ψ5A T T A T G T C
pHTPI-5a G
Leu
Intron 5
aa 170
aa 180
aa 190
Ala Ile Gly Thr Gly Lys Thr Ala Thr Pro Gln Gln Ala Gln Glu Val His Glu Lys Leu Arg Gly Trp Leu Lys Ser
8B GCC ATT GGT ACT GGC AAG ACT GCA ACA CCC CAA CAG GCC CAG GAA GTA CAC GAG AAG CTC CGA GGA TGG CTG AAG TCC
Ψ13C ***
Ψ19A C A G A T - ---
Ψ5A C C G A T
pHTPI-5a
Intron 6
aa 200
aa 210
aa 220
Asn Val Ser Asp Ala Val Ala Gln Ser Thr Arg Ile Ile Tyr Gly Gly Ser Val Thr Gly Ala Thr Cys Lys Glu Leu
8B AAC GTC TCT GAT GCG GTG GCT CAG AGC ACC CGT ATC ATT TAT GGA GGC TCT GTG ACT GGG GCA ACC TGC AAG GAG CTG
Ψ13C ***
Ψ19A A T T C TG
Ψ5A A A T G C AA GT --
pHTPI-5a


aa 230
aa 240
Ala Ser Gln Pro Asp Val Asp Gly Phe Leu Val Gly Gly Ala Ser Leu Lys Pro Glu Phe Val Asp Ile Ile Asn Ala Lys Gln Ter
8B GCC AGC CAG CCT GAT GTG GAT GGC TTC CTT GTG GGT GGT GCT TCC CTC AAG CCC GAA TTC GTG GAC ATC ATC AAT GCC AAA CAA TGA
Ψ13C ***
Ψ19A A C A A T
Ψ5A C C C C A A T A
pHTP1-5a
8B GCCCCATCCATCTTCCCTACCCTTCCTGCCAAGCCAGGGACTAAGCAGCCCAGAAGCCCAGTAACTGCCCTTTCCCTGCATATGCTTCTGATGGTGTCATCTGC
Ψ13C ************************************T- C C C
Ψ19A - - C C C T
Ψ5A G C A TG G C C C
pHTP1-5a
8B TCCTTCCTGTGGCCTCATCCAAACTGTATCTTCCTTTACTGTTTATATCTTCACCCTGTAATGGTTGGGACCAGGCCAATCCCTTCTCCACTTACTAT
Ψ13C C TC TT C A G
Ψ19A C C T C - C¶. ⟶
Ψ5A A C AT A G C G C A G
pHTP1-5a
8B AATGGTTGGAACTAAACGTCACCAAGGTGGCTTCTCCTTGGCTGAGAGATGGAAGGCGTGGTGGGATTTGCTCCTGGGTTCCCTAGGCCCTAGTGAGGGCAGAA
Ψ13C -AG-- C C G
Ψ5A T T T G A GA ---- A - C ¶ ⟶
pHTP1-5a
8B GAGAAACCATCCTCTCCCTTCTTACACCGTGAGGCCAAGATCCCCTCAGAAGGCAGGAGTGCTGCCCTCTCCCATGGTGCCCGTGCCTCTGTGCTGTGTATGTG
Ψ13C GT A CT A C A C C
pHTP1-5a
8B AACCACCCATGTGAGGGAATAAACCTAGCACTAGGTCTTGTGGTTTGTCTGCCTTCACTGGACTTGCCCAGATAATCTTCCTTTTTGAGGCAGCTATATAAATG
Ψ13C T A G AA G G AAAAAAAACACCTAAATTTGTTTTAAAATAACAAGGTTTATAAAATTATTGAATTTATTGGCCAATGAT
pHTP1-5a AAAAAAAAAAAAAA
8B ATCATTTGTGCAAGAAAAAAAAAAAAAACAAGAACAGGTTTCTATAACAACATCTCTTACTATTTTTACTTGAAAAAATGTTTTGCGTAGCAGACTGTCATAGCC
Ψ13C TTTGTGCAAAGGGGAAATGTGTAGTTCTTCGGTTAATTTTCATGTGTGACATGTGTACATCAACCACTCTTTAAACCCAGATTTCAAGAATTTCCCCTAAAAAT
8B TTGAACGCCGGCTCCCTTTCTTCCTCCCTCCAAGTGGCTCTGGGGCTGTTGATTTCCGCAGAGCTGGGTTGGGGTAGGGGCTCAGCCCACCAGCTTTCAG
Ψ13C GTTCCAAGGACCATGAGCTTAAAGTAAAAATAAGTC

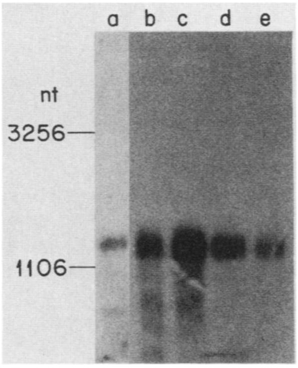


FIG. 3. Blot hybridization analysis of TPI RNA in cultured human cells. Poly(A)$^+$ RNA from 100 μg of total cellular RNA was electrophoresed in a 1.2% agarose–2.2 M formaldehyde slab gel, transferred to nitrocellulose, and hybridized with a ^{32}P-labeled subclone of TPI cDNA (27). The nitrocellulose was washed to remove unhybridized probe and exposed to X-ray film. Human L153 fibroblast RNA (a), HeLa cell RNA (b), Daudi Burkitt lymphoma B-cell RNA (c), Frawley Epstein-Barr virus-infected B-cell RNA (d), and Jurkat-FHCRC acute lymphocytic leukemia T-cell RNA (e). Restriction nuclease digests of pBR322 DNA were used as molecular weight standards. To detect these standards, the blot was washed to remove hybridized TPI probe and rehybridized with nick-translated pBR322 DNA.

TPI gene and, therefore, most probably the only functional TPI gene. Assuming that the uncharacterized TPI genes are processed pseudogenes of 1,225 bp, the four genomic clones described in this paper comprise approximately 65% of TPI gene sequences in the human genome.

DISCUSSION

We have characterized several members of the human TPI gene family. This family consists of a functional gene (hTPI-8B) and at least three processed pseudogenes (ΨhTPI-13C, ΨhTPI-19A, and ΨhTPI-5A). Each of these genes was isolated and analyzed in detail. The functional gene spans 3.5 kbp and is divided into seven exons. These exons encode a TPI polypeptide that is identical to that deduced from adult liver cDNA (27). The promoter includes sequences resembling TATA and CCAAT boxes and is unusually G-C rich (76%). The pseudogenes contain numerous bp substitutions and translational frameshift mutations that preclude the synthesis of a functional polypeptide. Each pseudogene bears the structural hallmarks of an origin from TPI cDNA that was integrated into germ line DNA via a transposition-like mechanism.

Evidence for a single functional TPI gene. Our data indicate that there is a single functional TPI gene per haploid genome.

TABLE 1. Exon-intron arrangement of the human TPI gene[a]

Intron no.	Exon-intron junction sequence (5′ junction.....3′ junction)		Intron size (bp)
1	CCGgtaagc..... accatctgtcctcag A nt 115 (Glu 38)	nt 116 (Glu 38)	~1,250
2	CAGgtgaga..... catctcttcctttag C nt 239 (Ser 79)	nt 240 (Ser 79)	111
3	GAGgttagt..... tctgtttctcaacag C nt 324 (Glu 107)	nt 325 (Leu 108)	74
4	CAGgtatct..... atctctgccctgcag A nt 457 (Asp 152)	nt 458 (Asp 152)	~310
5	CAGgtaacc..... agcttcttgttctag G nt 543 (Gln 180)	nt 544 (Ala 181)	~290
6	GAGgtgagt..... ttctgctccctccag G nt 631 (Gly 210)	nt 632 (Gly 210)	127
Consensus sequence	C/A AGgt a/g agt..... ttttttttttt/ccccccccccc n c/t agG		

[a] Exon sequences are in capital letters, and intron sequences are in lowercase letters. Introns 2, 3, and 6 were entirely sequenced. The sizes of introns 1, 4, and 5 are approximations. Nucleotide (nt) positions at which each intron interrupts TPI mRNA are numbered. The corresponding amino acids in TPI protein are shown within parentheses. When ambiguities in the precise splice site existed, exon-intron boundaries were determined by applying the gt-ag splice rule (38). Only intron nucleotides corresponding to the 5′ and 3′ consensus sequences (38) are shown.

All nine TPI cDNA sequences that were isolated from a human liver cDNA library are derived from the same mRNA species, suggesting that they were synthesized from transcripts of a single gene (27). A single-sized mRNA is detected by RNA blotting, and primer extension analysis together with pseudogene structure indicate that the 5′ terminus of this mRNA is localized to a single nucleotide. Only the functional hTPI-8B gene is detected when intron 4-specific sequences are hybridized to restriction enzyme digests of human DNA, suggesting that the remaining TPI genomic sequences are intronless pseudogenes. In agreement with this hypothesis, most *Eco*RI and *Pst*I TPI genomic fragments hybridize to both the 5′ and 3′ halves of TPI cDNA (data not shown), as expected for processed TPI genes which should consist of approximately 1,225 bp.

FIG. 2. Sequence of the functional TPI gene and three TPI processed pseudogenes. The sequences of three TPI processed pseudogenes (ΨhTPI-13C, ΨhTPI-19A, and ΨhTPI-5A) and the cDNA from pHTPI-5a (27) are compared to the exon sequences of the functional hTPI-8B gene. Only pseudogene and cDNA nucleotides that differ from functional gene nucleotides are shown. Genomic DNAs were isolated from a single individual (26). The cDNA was derived from an unrelated individual (60). Nucleotides upstream from the coding region of hTPI-8B are numbered relative to the first nucleotide of the translation initiation codon (+1). Intron positions in hTPI-8B are indicated. The putative TATA box, CCAAT box, and AATAAA polyadenylation-transcript cleavage site in hTPI-8B are underlined. The transcription start site is represented by a broken arrow (⌈--→). Deletions (−) and insertions (⋏) have been introduced into the ΨhTPI-13C, ΨhTPI-19A, and ΨhTPI-5A genes to retain maximum homology with the functional gene. Asterisks in the ΨhTPI-13C sequence designate the 317-bp deletion that is thought to be the result of a homologous recombination event. The exact crossover site is unknown; therefore, the deletion is shown to include the entire first repeat and none of the second repeat. Direct repeats involved in the generation of this deletion are underlined twice in hTPI-8B. Short, direct repeats flanking the ΨhTPI-13C gene are within boxes. ¶, Not determined.

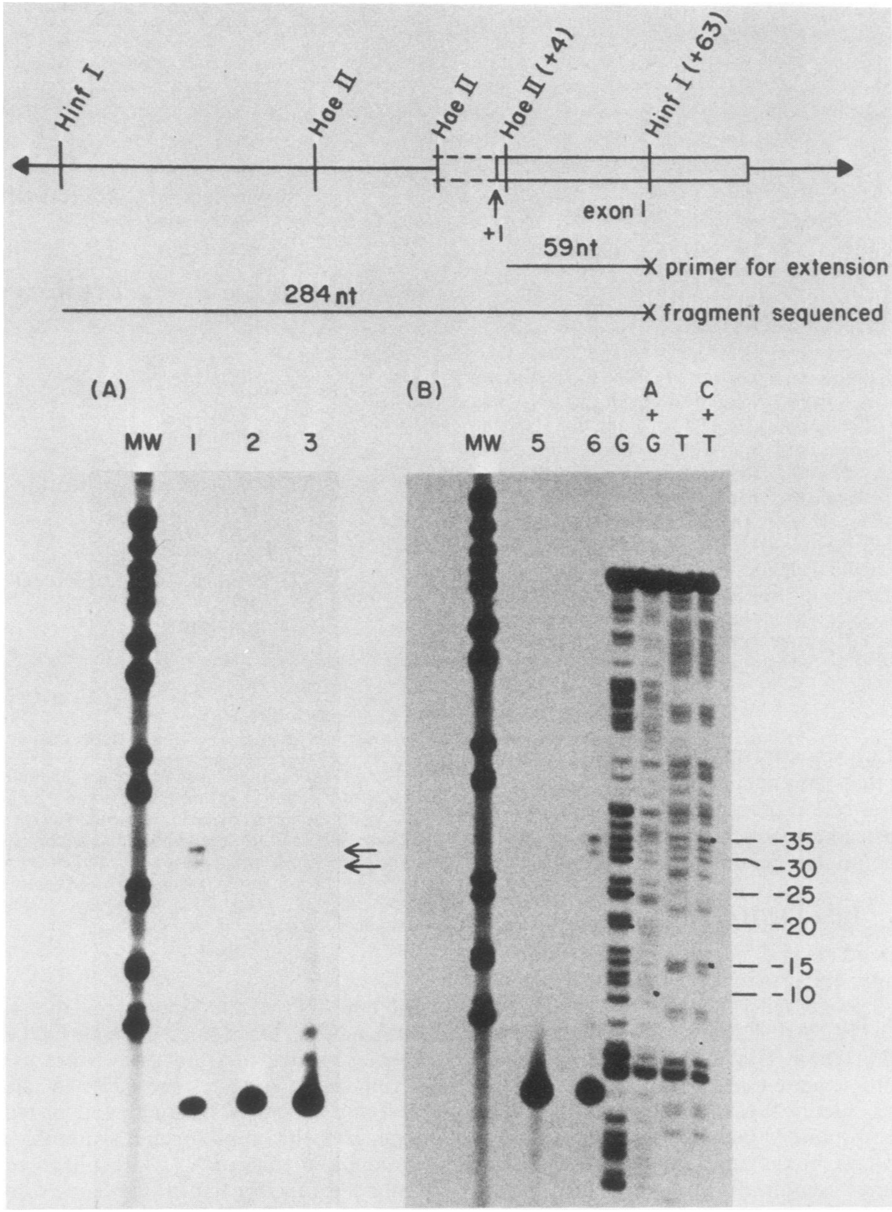


FIG. 4. Mapping the 5′ end of TPI mRNA by primer extension. Poly(A)$^+$ RNA (1 μg) from Frawley Epstein-Barr virus-infected B cells was hybridized with 25 ng of a single-stranded, 5′-^{32}P-labeled *Hin*fI-*Hae*II fragment (1.6 × 10^3 cpm/ng) complementary to nucleotides +4 to +63 of the hTPI-8B genomic sequence (where +1 is the first nucleotide of the translation initiation codon). The primer was extended, and extension products were characterized by electrophoresis in an 8% polyacrylamide–7 M urea gel (see the text). (A) MW, *Hpa*II-digested pBR322 DNA, 3′ end labeled with ^{32}P; lane 1, primer-extended poly(A)$^+$ RNA; lane 2, primer alone; lane 3, primer extended in the absence of poly(A)$^+$ RNA. Arrows indicate the two primer extension products. (B) MW, *Hpa*II-digested pBR322 DNA, 3′ end labeled with ^{32}P; lane 5, primer alone; lane 6, primer-extended poly(A)$^+$ RNA; lanes G, A+G, T, C+T, Maxam-Gilbert sequencing reactions. The 5′-^{32}P-labeled terminus of the sequenced fragment corresponds to that of the primer used in the extensions. Since the sequencing reactions result in the release of the modified (3′-terminal) base, a single nucleotide is added to each sequencing product to denote the actual distance from the ^{32}P-labeled end (e.g., the first nucleotide of the translation initiation codon, normally designated +1, is assigned position +2).

Previous gene mapping studies by TPI enzyme activity determinations placed the expressed gene on chromosome 12 (18, 20, 23). Using human-rodent hybrid cell DNAs, we have found that the functional hTPI-8B gene maps to this chromosome and that the processed pseudogenes are dispersed to other chromosomes (R. Eddy, J. R. Brown, and L. E. Maquat, unpublished data). Our data supporting the existence of only one expressed TPI gene concur with data of previous TPI isozyme studies. These studies indicate that the multiple electrophoretic and chromatographic forms of this enzyme differ only by the degree of posttranslational modification (12, 15, 42, 51).

Evolutionary history of TPI pseudogenes. The evolutionary relationship between a pseudogene and its functional counterpart can be estimated by calculating the sequence divergence between the protein-coding regions of the two genes (14, 41). Sequence divergence is often measured in terms of nucleotide substitutions that either do not lead to amino acid

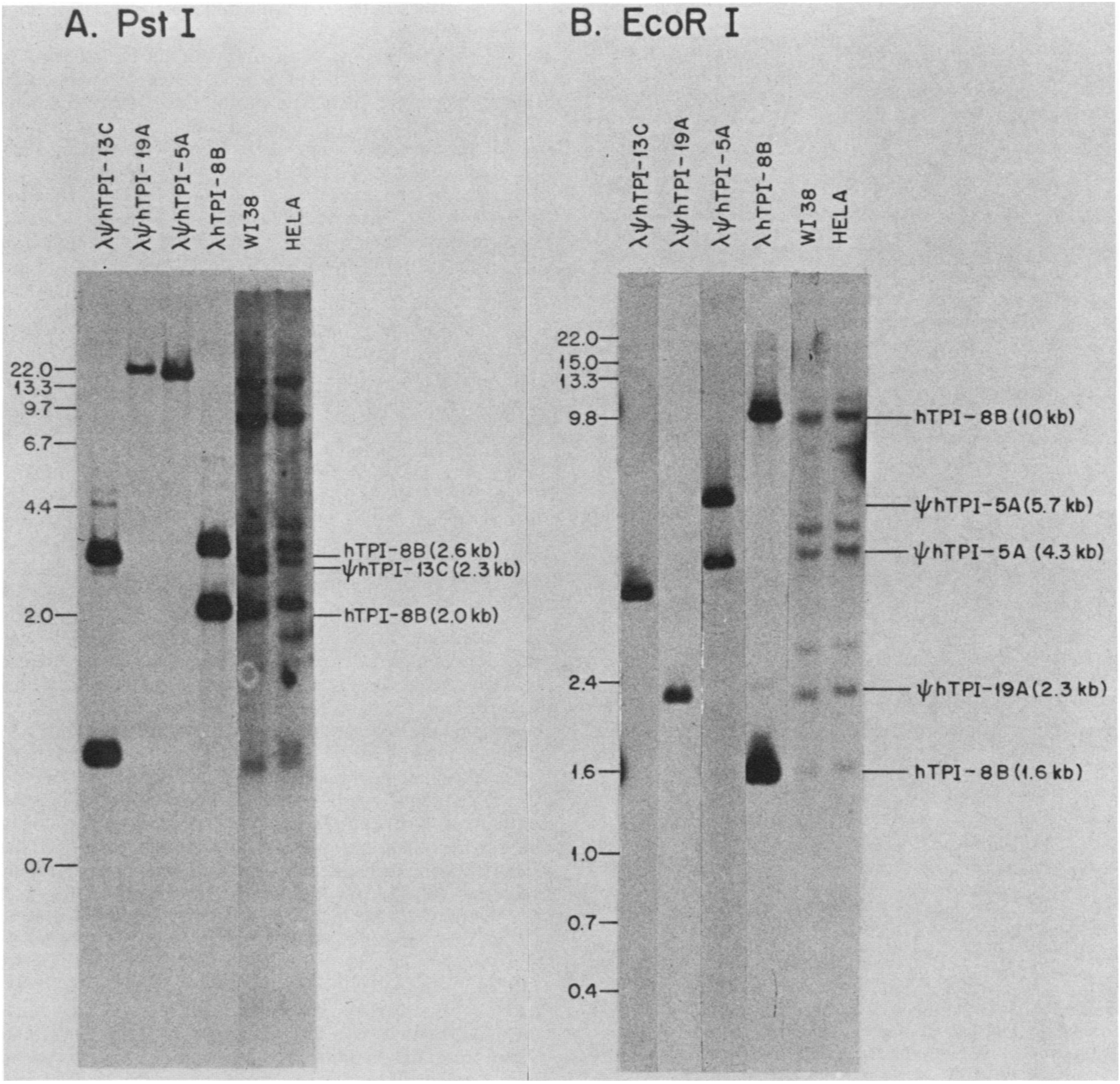


FIG. 5. TPI sequences in genomic DNA. Genomic DNA (10 μg) or recombinant bacteriophage DNA (2 ng) was digested with a restriction enzyme, electrophoresed in a 0.8% agarose gel, transferred to nitrocellulose, and hybridized with a nick-translated cDNA insert from pHTPI-5a. Filters were washed to remove unhybridized probe and exposed to X-ray film. *Pst*I (A) and *Eco*RI (B) restriction fragments of HeLa or WI38 genomic DNA that comigrate with TPI sequences in recombinant bacteriophage DNA are indicated. Restriction fragments were sized relative to a *Hin*dIII digest of λ DNA.

replacements (silent site substitutions) or do lead to amino acid replacements (replacement site substitutions).

The numbers of nucleotide substitutions per site for each of the TPI pseudogenes relative to the functional gene were determined by the method of Perler et al. (41) (Table 2). Considering that the replacement sites of processed pseudogenes have not been under selective pressure for the entire time, if for any time, since the pseudogenes arose from the functional gene, silent site substitutions more accurately predict pseudogene divergence times (22). Accordingly, divergence times for each of the TPI pseudogenes were calculated by using the silent site substitution rates established for globin genes (14). These rates should be applicable to nonglobin genes since, aside from mRNA structural constraints (e.g., constraints dictated by mRNA secondary structure or specific codon requirements), the silent sites of any gene should not be under selective pressure (8, 41). Silent site substitutions predict that the TPI pseudogenes began to diverge from the functional gene approximately 18 million years ago (Table 2).

Divergence of the 34-bp 5′ untranslated region of each of the pseudogenes from the functional gene (Table 2) is in part due to a 3-bp sequence that is present in the pseudogenes yet lacking in the hTPI-8B gene (Fig. 2). It is likely that the three

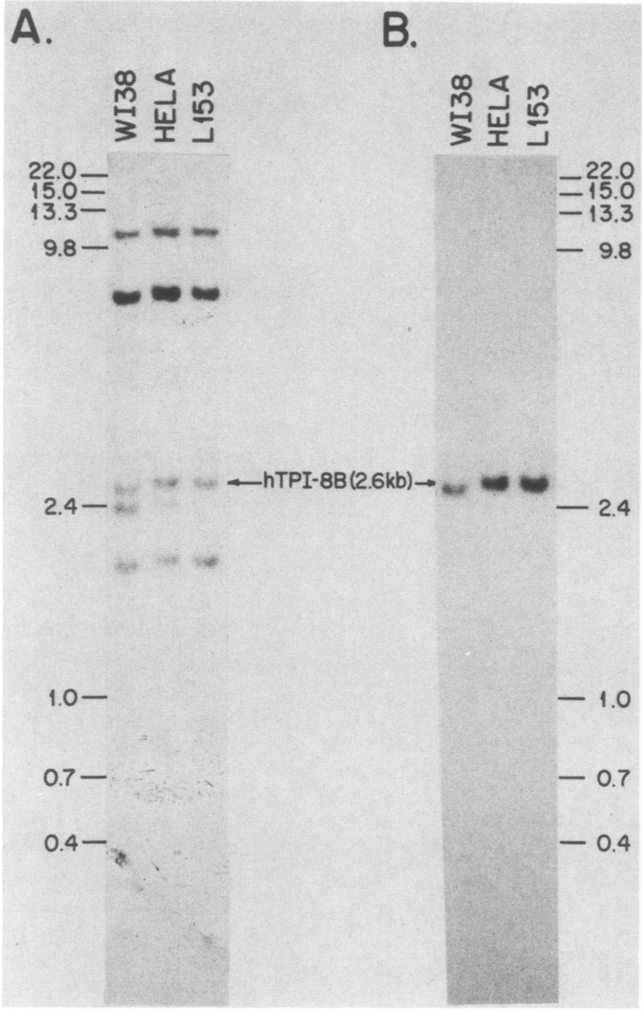


FIG. 6. hTPI-8B intron 4-homologous sequences in genomic DNA. Genomic DNA (10 μg) was digested with *Pst*I, electrophoresed in a 0.8% agarose gel, transferred to nitrocellulose, and hybridized with a [^{32}P]cDNA insert from pHTPI-5a (A). After exposure to X-ray film, the nitrocellulose was washed to remove hybridized cDNA and rehybridized to ^{32}P-labeled intron 4 sequences from the hTPI-8B gene (B). Note that these DNA samples were electrophoresed for a longer time than those shown in Fig. 5. Therefore, some of the smaller TPI-homologous restriction fragments are less apparent in this figure than in Fig. 5.

nucleotides were a part of the 5' untranslated region of the functional gene when the pseudogene arose. Similarly, since the pseudogenes as well as the pHTPI-5a cDNA contain a G whereas the hTPI-8B gene contains a C at the third position of amino acid codon 162, it is likely that the hTPI-8B gene at one time also contained a G at this position. Nucleotides in common to two or more pseudogenes and not present in the hTPI-8B gene are evident throughout the coding and untranslated regions. At least some of these undoubtedly reflect changes in the functional gene since the time of pseudogene divergence.

Since TPI genomic sequences were isolated from the recombinant library by relatively stringent hybridization (5× SSC, 50% formamide, 37°C) and wash (0.1× SSC, 0.1% sodium dodecyl sulfate, 50°C) conditions, we presume that we have characterized the most conserved members of this gene family. Processed TPI pseudogenes that evolved before those described in this paper most likely also exist within the human genome.

The hTPI-8B gene was also compared to the *Saccharomyces cerevisiae* TPI gene (1). There have been multiple changes per silent site during the estimated 1 billion years (37) since the evolutionary branch that gave rise to yeasts split from the branch that gave rise to humans (Table 2). Replacement site rather than silent site substitutions more accurately reflect the divergence times of functional genes (14, 41). However, the lack of TPI gene sequences from other organisms more closely related to humans precludes the establishment of a clock based on TPI gene evolutionary rates. Calculations of TPI gene divergence times based, for example, on globin gene evolutionary rates (14) would be inappropriate since genes encoding proteins with different structural constraints are known to have different rates of replacement site changes (41, 59). A search for sequence homologies within the 5' and 3' untranslated regions of the human and yeast genes revealed no conserved remnant of a common ancestral gene other than the transcript cleavage-polyadenylation signal (data not shown).

The TPI gene promoter. In searching for elements that govern housekeeping gene expression in all cells, one might take a reductionist approach and propose that these elements (i) act at the level of gene transcription, (ii) should be shared by all housekeeping genes, and (iii) should be absent in facultative genes whose expression is limited to a particular developmental stage. *cis*-Acting transcriptional regulators of polymerase II-transcribed genes have been localized to DNA sequences both upstream (7, 11, 58) and downstream (3, 9, 61) from the transcription start site. Since the upstream control elements are better characterized, we compared the putative TPI gene promoter with promoters of other housekeeping genes. From this comparison, we were unable to identify any unifying feature of housekeeping gene promoters. Although housekeeping gene promoters for mouse hypoxanthine phosphoribosyltransferase (32), hamster HMG coenzyme A reductase (44), and mouse ribosomal protein L30 (57) lack TATA (4, 16) and CCAAT (5, 14) boxes, the TPI promoter has both canonical TATA and CCAAT sequences. The TPI promoter is characterized by an extremely high G-C content. DNA sequences extending 140 bp upstream from the start of transcription are 76% G and C residues with a preponderance (58%) of G residues. A high G-C content is a feature of some, but not all, housekeeping gene promoters. The mouse hypoxanthine phosphoribosyltransferase and hamster HMG coenzyme A reductase genes have promoter regions that are 77 and 65% G+C, respectively (32, 44); however, the rat cytochrome *c* gene promoter, for example, is only 38% G+C (48).

Housekeeping genes are similar to immediate early viral genes in that they are recognized by the transcriptional machinery of cells in the absence of any modifications to this machinery. It follows that the transcriptional control regions of housekeeping genes and immediate early viral genes might be homologous. However, the TPI sequence lacks significant homology to any viral enhancer (58) or to putative enhancers in human DNA that have been identified by homology to the simian virus 40 72-bp repeat (10).

Although the herpes simplex virus (HSV) tk gene requires induction by early viral gene products (43), its promoter (29) shares certain sequences with the TPI promoter that merit mention. These sequences include the GTGGCC at approximately position −10 and the CCACTTCGC at approximately position −35 (where +1 is the transcription start site). However, HSV tk gene mutagenesis failed to reveal a

TABLE 2. Divergence of TPI gene sequences[a]

Gene pair	Replacement site divergence of coding region +1 to +747 (nucleotide changes per site)	Silent site divergence at: Coding region +1 to +747		5′ untranslated region −34 to −1 (nucleotide changes per site)	3′ untranslated region +748 to +1195 (nucleotide changes per site)
		Nucleotide changes per site	Divergence time (myr)		
ψhTPI-13C vs. hTPI-8B	0.081	0.15	19	0.191	0.114
ψhTPI-19A vs. hTPI-8B	0.075	0.14	18	0.191	ND
ψTPI-5A vs. hTPI-8B	0.085	0.14	18	0.229	0.115
Yeast TPI vs. hTPI-8B	0.435	2.10		ND	ND

[a] The nucleotide changes per site within coding regions were determined by the method of Perler et al. (41). Times of divergence for each TPI pseudogene were calculated assuming 0.008 substitutions per silent site per 1 million years for two initially identical sequences (14, 41). When comparing untranslated regions, a nucleotide in common to the two regions was scored as +1, a substitution, deletion, or insertion was scored as −1, and the denominator (total number of nucleotides evaluated) was taken as the average number of nucleotides in the two sequences under comparison. Nucleotide changes per site within the untranslated regions were corrected for multiple events by the formula of Jukes and Cantor (19): $d = (3/4)\ln[1-(4/3)p]$, where d equals the corrected number and p equals the uncorrected number of nucleotide changes per site. A number larger than unity reflects corrections for multiple changes at a single site (41). ND, Not determined.

functional role for either of these sequences (30, 31). Homologies to the GTGGCC sequence of the HSV tk gene are also present in the chicken tk (33) and the mouse hypoxanthine phosphoribosyltransferase (32) genes at similar positions. The TPI gene also shares considerable homology to the first distal transcriptional signal of the HSV tk gene (30, 31). Of the 14 bp shown to be functionally important for HSV tk gene transcription, 9 are present within the TPI promoter at the same location (bp −61 to −48). Proof for the significance of these similarities and the possible existence of regulatory sequences located downstream from the start site of TPI gene transcription requires further studies.

ACKNOWLEDGMENTS

We thank Patricia Ryan for expert technical assistance, Tom Maniatis for the human genomic library, Loren Field for help with primer extensions, Argiris Efstratiadis for the nucleotide sequence divergence program, Ken Manly for adapting this program, Ken Gross and Dave Kowalski for the use of computer facilities, Roger Eddy and Tom Shows for human-mouse cell hybrid DNA, John Pauley for B- and T-cell lines, Ross Hardison, Alan Kinniburgh, and Ken Gross for helpful discussions, and Ina Young for excellent secretarial assistance.

This work was supported by Public Health Service grants RO1 AM 31747 and AM/GM 33938 from the National Institutes of Health. L.E.M. is an Established Investigator of the American Heart Association.

LITERATURE CITED

1. **Alber, T., and G. Kawasaki.** 1982. Nucleotide sequence of the triosephosphate isomerase gene of *Saccharomyces cerevisiae*. J. Mol. Appl. Genet. **1**:419–434
2. **Aviv, H., and P. Leder.** 1972. Purification of biologically active globin messenger RNA by chromatography on oligothymidylic acid-cellulose. Proc. Natl. Acad. Sci. U.S.A. **69**:1408–1412.
3. **Banerji, J., L. Olson, and W. Shaffner.** 1983. A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes. Cell **33**:729–740.
4. **Benoist, C., and P. Chambon.** 1981. *In vivo* sequence requirements of the SV40 early promoter region. Nature (London) **290**:304–310.
5. **Benoist, C., K. O'Hare, R. Breathnach, and P. Chambon.** 1980. The ovalbumin gene sequence of putative control regions. Nucleic Acids Res. **8**:127–142.
6. **Benton, W. D., and R. W. Davis.** 1977. Screening gt recombinant clones by hybridization to single plaques *in situ*. Science **196**:180–182.
7. **Breathnach, R., and P. Chambon.** 1981. Organization and expression of eucaryotic split genes coding for proteins. Annu. Rev. Biochem. **50**:349–383.
8. **Chan, S. J., V. Episkopou, S. Zeitlin, S. K. Karathananasis, A. MacKrell, D. F. Steiner, and A. Efstratiadis.** 1984. Guinea pig preproinsulin gene: an evolutionary compromise? Proc. Natl. Acad. Sci. U.S.A. **81**:5046–5050.
9. **Charnay, P., R. Treisman, P. Mellon, M. Chao, R. Azel, and T. Maniatis.** 1984. Differences in human α- and β-globin gene expression in mouse erythroleukemia cells: the role of intragenic sequences. Cell **38**:251–263.
10. **Conrad, S. E., and M. R. Botchan.** 1982. Isolation and characterization of human DNA fragments with nucleotide sequence homologies with the simian virus 40 regulatory region. Mol. Cell. Biol. **2**:949–965.
11. **Corden, J., B. Wasylyk, A. Buchwalder, P. Sassone-Corsi, C. Kedinger, and P. Chambon.** 1980. Expression of cloned genes in new environments. Science **209**:1406–1414.
12. **Decker, R. S., and H. W. Mohrenweiser.** 1981. Origin of the triosephosphate isomerase isozymes in humans; genetic evidence of the expression of a single structural locus. Am. J. Hum. Genet. **33**:683–691.
13. **Eber, S. W., M. Dunnwald, B. H. Belohradsky, F. Bidlingmaier, H. Schievelbein, H. M. Weinmann, and K. G. Kreitsch.** 1979. Hereditary deficiency of triosephosphate isomerase in four unrelated families. Eur. J. Clin. Invest. **9**:195–202.
14. **Efstratiadis, A., J. Posakony, T. Maniatis, R. Lawn, C. O'Connell, R. Spritz, J. DeReil, B. Forget, S. Weissman, J. Slightom, A. Blechl, D. Smithies, F. Baralle, C. Shoulders, and N. Proudfoot.** 1980. The structure and evolution of the human β-globin gene family. Cell **21**:653–668.
15. **Gracy, R. W.** 1982. Glucosephosphate and triosephosphate isomerases: significance of isozyme structural differences in evolution, physiology and aging. Isozymes Curr. Top. Biol. Med. Res. **6**:169–205.
16. **Grosschedl, R., and M. L. Birnstiel.** 1980. Identification of regulatory sequences in the prelude sequences of an HLA histone gene by the study of specific deletion mutants *in vivo*. Proc. Natl. Acad. Sci. U.S.A. **77**:1432–1436.
17. **Hollis, G. F., P. A. Hieter, O. W. McBride, D. Swan, and P. Leder.** 1982. Processed genes: a dispersed human immunoglobulin gene bearing evidence of RNA-type processing. Nature (London) **296**:321–325.
18. **Jongsma, A. P. M., A. Hagemeijer, and P. M. Khan.** 1975. Regional mapping of TPI, LDH-B, and Pep-B on chromosome 12 in man. Birth Defects **1163**:189–191.
19. **Jukes, T. H., and C. R. Cantor.** 1969. Mammalian protein metabolism, p. 21–123. Academic Press, Inc., New York.
20. **Junien, C., J. C. Kaplan, F. Serville, and G. Lenoir.** 1979. Triplex gene dosage effects of TPI and G3PD in a human lymphoblastoid cell line with partial trisomy 12p13 and 18p. Hum. Genet. **49**:221–223.
21. **Kaplan, J. C., L. Teeple, N. Shore, and E. Beutler.** 1968.

Electrophoretic abnormality in triosephosphate isomerase deficiency. Biochem. Biophys. Res. Commun. 31:768–773.
22. Lacy, E., and Maniatis, T. 1980. The nucleotide sequence of a rabbit β-globin pseudogene. Cell 21:545–553.
23. Law, M. L., and F. T. Kao. 1979. Regional assignment of human genes TPI, GAPDH, LDH_B, SHMT, and PEPB on chromosome 12. Cytogenet. Cell Genet. 24:102–114.
24. Lu, H. S., P. M. Yuan, and R. W. Gracy. 1984. Primary structure of human triosephosphate isomerase. J. Biol. Chem. 259:11958–11968.
25. Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular cloning, a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
26. Maniatis, T., R. C. Hardison, E. Lacy, J. Lauer, C. O'Connell, D. Quon, D. K. Sim, and A. Efstratiadis. 1978. The isolation of structural genes from libraries of eucaryotic DNA. Cell 15:687–701.
27. Maquat, L. E., Chilcote, R., and P. M. Ryan. 1985. Human triosephosphate isomerase cDNA and protein structure: studies of triosephosphate isomerase deficiency in man. J. Biol. Chem. 260:3748–3753.
28. Maxam, A. M., and W. Gilbert. 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. 65:499–560.
29. McKnight, S. L. 1980. The nucleotide sequence and transcript map of the herpes simplex virus thymidine kinase gene. Nucleic Acids Res. 8:5949–4964.
30. McKnight, S. L., and R. Kingsbury. 1982. Transcriptional control signals of a eukaryotic protein coding gene. Science 217:316–324.
31. McKnight, S. L., R. C. Kingsbury, A. Spence, and M. Smith. 1984. The distal transcription signals of the herpes virus tk gene share a common hexanucleotide control sequence. Cell 37:253–262.
32. Melton, D. W., D. S. Konecki, J. Brennand, and C. T. Caskey. 1984. Structure, expression, and mutation of the hypoxanthine phosphoribosyltransferase gene. Proc. Natl. Acad. Sci. U.S.A. 81:2147–2151.
33. Merrill, G. F., R. M. Harland, M. Groudine, and S. L. McKnight. 1984. Genetic and physical analysis of the chicken *tk* gene. Mol. Cell. Biol. 4:1769–1776.
34. Mohrenweiser, H. W. 1981. Frequency of enzyme deficiency variants in erythrocytes of newborn infants. 1981. Proc. Natl. Acad. Sci. U.S.A. 78:5046–5050.
35. Mohrenweiser, H. W., and S. Fielek. 1982. Elevated frequency of carriers for triosephosphate isomerase deficiency in newborn infants. Pediatr. Res. 16:960–963.
36. Mohrenweiser, H. W., S. Fielek, and K. H. Wurzinger. Characteristics of enzymes of erythrocytes from newborn infants and adults. Activity, thermostability and electrophoretic profile as a function of cell age. Am. J. Hematol. 11:125–136.
37. Moore, G. W., M. Goodman, C. Callahan, R. Holmquist, and H. Moise. 1976. Stochastic versus augmented maximum parsimony method for estimating superimposed mutations in the divergent evolution of protein sequences. Methods tested on cytochrome c amino acid sequences. J. Mol. Biol. 105:15–37.
38. Mount, S. M. 1982. A catalogue of splice junction sequences. Nucleic Acids Res. 10:459–472.
39. Nishioka, Y., A. Leder, and P. Leder. 1980. Unusual α-globin-like gene that has cleanly lost both globin intervening sequences. Proc. Natl. Acad. Sci. U.S.A. 77:2806–2809.
40. Noltmann, E. A. 1972. Aldose ketose isomerases. Enzymes 1:271–353.
41. Perler, F., A. Efstratiadis, P. Lomedico, W. Gilbert, R. Kolodner, and J. Dodgson. 1980. The evolution of genes: the chicken preproinsulin gene. Cell 20:555–566.
42. Peters, J., D. A. Hopkinson, and H. Harris. 1973. Genetic and non-genetic variation of triosephosphate isomerase isozymes of human tissues. Ann. Hum. Genet. 36:297–312.
43. Post, L. E., S. Mackem, and B. Roizman. 1981. The regulation of α genes of herpes simplex virus: expression of chimeric genes produced by fusion of thymidine kinase with α gene promoter. Cell 24:555–565.
44. Reynolds, G. A., S. K. Basu, T. F. Osborne, D. J. Chin, G. Gil, M. S. Brown, J. L. Goldstein, and K. L. Luskey. 1984. HMG CoA reductase: a negatively regulated gene with an unusual promoter and 5′ untranslated region. Cell 38:275–285.
45. Ross, J. 1976. A precursor of globin messenger RNA. J. Mol. Biol. 106:403–420.
46. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. U.S.A. 74:5463–5467.
47. Satoh, C., J. V. Neel, A. Yamashita, K. Goriki, M. Fujita, and H. H. Hamilton. 1983. The frequency among Japanese of heterozygotes for deficiency variants of 11 enzymes. Am. J. Hum. Genet. 35:656–674.
48. Scarpulla, R. C., and R. Wu. 1983. Nonallelic members of the cytochrome c multigene family of the rat may arise through different messenger RNAs. Cell 32:473–482.
49. Schneider, A. S., W. N. Valentine, M. Hattori, and H. L. Heins. 1965. Hereditary hemolytic anemia with triosephosphate isomerase deficiency. N. Engl. J. Med. 272:229–235.
50. Skala, H., J. C. Dreyfus, J. L. Vives-Corrons, F. Matsumoto, and E. Beutler. 1977. Triose phosphate isomerase deficiency. Biochem. Med. 18:226–234.
51. Snapka, R. M., T. H. Sawyer, R. A. Barton, and R. W. Gracy. 1974. Comparison of the electrophoretic properties of triosephosphate isomerases of various tissues and species. Comp. Biochem. Physiol. 49B:733–744.
52. Southern, E. M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98:503–517.
53. Thomas, P. S. 1980. Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. Proc. Natl. Acad. Sci. U.S.A. 77:5201–5205.
54. Valentine, W. N., K. R. Tanaka, and D. E. Paglia. 1983. Pyruvate kinase and other enzyme deficiency disorders of the erythrocyte, p. 1606–1628. *In* J. B. Stanbury, J. B. Wyngaarden, D. S. Frederickson, J. L. Goldstein, and M. S. Brown (ed.), The metabolic basis of inherited disease. McGraw-Hill Book Co., New York.
55. Vanin, E. F., G. I. Goldberg, P. W. Tucker, and O. Smithies. 1980. A mouse α-globin-related pseudogene lacking intervening sequences. Nature (London) 286:222–226.
56. Vives-Corrons, J.-L., H. Robinson-Skala, M. Inateo, J. Estella, E. Feliu, and J. C. Dreyfus. 1978. Triosephosphate isomerase deficiency with hemolytic anemia and severe neuromuscular disease. Familial and biochemical studies of a case found in Spain. Hum. Genet. 42:171–180.
57. Wiedemann, L. M., and R. P. Perry. 1984. Characterization of the expressed gene and several processed pseudogenes for the mouse ribosomal protein L30 gene family. Mol. Cell. Biol. 4:2518–2528.
58. Weiher, H., M. Konig, and P. Gross. 1983. Multiple point mutations affecting the simian virus 40 enhancer. Science 219:616–631.
59. Wilson, A. C., S. S. Carlson, and T. J. White. 1977. Biochemical evolution. Annu. Rev. Biochem. 46:573–639.
60. Woods, D. E., A. F. Markam, A. T. Ricker, G. Goldberger, and H. R. Colten. 1982. Isolation of cDNA clones for the human complement protein factor B, a class III major histocompatibility complex gene product. Proc. Natl. Acad. Sci. U.S.A. 79:5661–5665.
61. Wright, S., A. Rosenthal, R. Flavell, and F. Grosveld. 1984. DNA sequences required for regulated expression of β-globin genes in murine erythroleukemia cells. Cell 38:265–273.