ANALYSIS OF A HUMAN TRANSFER RNA GENE CLUSTER AND CHARACTERIZATION OF THE TRANSCRIPTION UNIT AND TWO PROCESSED PSEUDOGENES OF CHIMPANZEE TRIOSEPHOSPHATE ISOMERASE

DISSERTATION

Presented to the Graduate Council of the
University of North Texas in Partial
Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

By

Leonard C. Craig, B.S., M.B.A., B.S.
Denton, Texas

August, 1990

Craig, Leonard C., Analysis of a Human Transfer RNA Gene Cluster and Characterization of the Transcription Unit and Two Processed Pseudogenes of Chimpanzee Triosephosphate Isomerase. Doctor of Philosophy (Biochemistry), August, 1990, 109 pp., 9 illustrations, bibliography, 155 titles.

An 18.5-kb human DNA segment was selected from a human λCharon-4A library by hybridization to mammalian valine $tRNA_{IAC}$ and found to encompass a cluster of three tRNA genes. Two valine tRNA genes with anticodons of AAC and CAC, encoding the major and minor cytoplasmic valine tRNA isoacceptors, respectively, and a lysine $tRNA_{CUU}$ gene were identified by Southern blot hybridization and DNA sequence analysis of a 7.1-kb region of the human DNA insert. At least nine *Alu* family members were found interspersed throughout the human DNA fragment. The tRNA genes are accurately transcribed by RNA polymerase III in a HeLa cell extract, since the RNase T_1 fingerprints of the mature-sized tRNA transcription products are consistent with the DNA sequences of the structural genes.

Three members of the chimpanzee triosephosphate isomerase (TPI) gene family, the functional transcription unit and two processed pseudogenes, were characterized by genomic blotting and DNA sequence analysis. The *bona fide* TPI gene spans 3.5 kb with seven exons and six introns, and

is the first complete hominoid TPI gene sequenced. The gene exhibits a very high identity with the human and rhesus TPI genes. In particular, the polypeptides of 248 amino acids encoded by the chimpanzee and human TPI genes are identical, although the two coding regions differ in the third codon wobble positions for five amino acids. An *Alu* member occurs upstream from one of the processed pseudogenes, whereas an isolated endogenous retroviral long terminal repeat (HERV-K) occurs within the structural region of the other processed pseudogene. The ages of the processed pseudogenes were estimated to be 2.6 and 10.4 million years, implying that one was inserted into the genome before the divergence of the chimpanzee and human lineages, and the other inserted into the chimpanzee genome after the divergence.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

Figure Page

CHAPTER 1

ANALYSIS OF A HUMAN TRANSFER RNA GENE CLUSTER

Introduction

Each of the 60-90 different isoaccepting species of tRNA (Lin and Agris, 1980) appears to be encoded by a human tRNA gene family of about 12 members, since there are about 1300 tRNA genes in the human genome (Hatlen and Attardi, 1971). However, this has only been demonstrated for two tRNA gene families, and only three or four members of each family have been studied to date. Twelve human initiator methionine tRNA genes occur individually and are dispersed throughout the human genome (Santos and Zasloff, 1981). A minimum of 13 valine tRNA genes were detected in genomic blots of human DNA (Arnold *et al.*, 1986). Thus, determination of the structures and genomic organization of all members of a tRNA multigene family should lead to more definitive correlations between tRNA gene structure and regulation of expression within a human isoacceptor tRNA multigene family.

The *Alu* family of short interspersed repeated sequences (SINEs) consists of 500,000 copies in the human genome (Weiner *et al.*, 1986). This primate retropseudogene

is about 280 bp long, and is composed of two imperfect, tandemly repeated, 140-bp monomer units containing a variable oligo(dA) tract at the 3' end and flanked by short direct repeat sequences representing the duplicated target sites for retroposition. A number of *Alu* elements are efficiently transcribed by RNA polymerase III due to the presence of an internal split promoter (Elder *et al.*, 1981). Thus, *Alu* members could have profound influences on human tRNA gene expression, especially if their sites of integration are near tRNA gene clusters. Rosen and Daniel (1988) showed that a member of the rodent B2 SINE family neighboring a rat phenylalanine tRNA gene was transcribed more efficiently than the tRNA gene, because the retroposon could form a more stable transcription complex than the tRNA gene and could compete with it for binding of transcription factors. *Alu* elements have only been identified near a few human tRNA genes (Arnold *et al.*, 1986; Doran *et al.*, 1987; Shortridge *et al.*, 1989). Doran *et al.*(1987) found at least eight putative *Alu* members within a human DNA segment of 13.8 kb encompassing a cluster of four tRNA genes, whereas Arnold *et al.* (1986) found *Alu* members within several hundred base pairs of two solitary valine tRNA genes. Thus, it is possible that nonrandom integration of *Alu* members near tRNA genes is a general feature of tRNA gene loci within the human genome. This possibility remains conjectural until more extensive sequence information about

human DNA segments encoding tRNA genes is available.

Arnold *et al.* (1986) characterized three solitary members of the human valine tRNA isoacceptor gene family, and later demonstrated (Arnold and Gross, 1987; Arnold *et al.*, 1988) that both flanking regions of a valine $tRNA_{AAC}$ gene contain extragenic control regions that serve as positive modulatory elements for transcription of the valine tRNA gene. Another member of the human valine tRNA multigene family in a heterocluster of three tRNA genes has also been characterized (Shortridge *et al.*, 1989). Kahnt *et al.* (1989) analyzed a variant valine $tRNA_{CAC}$ gene that differs by four base substitutions from a functional valine $tRNA_{CAC}$ gene (Arnold *et al.*, 1986), and which is transcribed efficiently in a homologous HeLa cell extract.

Since it is important to obtain further information on the genomic organization and expression of one tRNA isoacceptor gene family, this work continues the studies of the human valine tRNA gene family by characterizing a human DNA segment encompassing a tRNA gene cluster, encoding both the major and minor valine tRNA isoacceptors and a lysine $tRNA_{CUU}$ isoacceptor, and also encompassing at least nine *Alu* family members (Craig *et al.*, 1989).

Materials and Methods

Genomic Library Screening and Physical Mapping

Bovine liver valine [3'-^{32}P]tRNA$_{IAC}$, purified by Irma Pirtle and Robert Pirtle (Johnson *et al.*, 1985), was used as a hybridization probe to screen a human gene library harbored in bacteriophage λ Charon-4A (Slightom *et al.*, 1980), as described (Pirtle *et al.*, 1986). One recombinant phage that gave an intense positive signal, subsequently designated as λhVKV7, was plaque-purified and characterized by Southern blot hybridization analyses using unfractionated bovine liver [3'-^{32}P]tRNA. The physical map of λhVKV7 was constructed primarily by Luping Wang. To construct the physical map of λhVKV7 and select appropriate restriction fragments for subcloning, single and double digestions of the λhVKV7 DNA were done using restriction endonucleases from New England BioLabs and Bethesda Research Laboratories (BRL). The digests were fractionated on 0.7%, 1.0% or 2.0% agarose gels, stained with ethidium bromide and photographed. To separate and determine the sizes of restriction fragments less than 1.0-kb, 4% agarose gels (NuSieve from Marine Colloids) were used. Denatured DNAs from some gels were transferred to nitrocellulose (Southern, 1979) in 0.3 M sodium citrate, 3.0 M NaCl (20x SSC) while partially depurinated and denatured DNAs from other gels

were transferred to positively-charged nylon membranes (Zeta-Probe from BioRad) in 0.4N NaOH (Reed and Mann, 1985) before hybridization to the ^{32}P-labeled probes. Unfractionated bovine liver [3'-^{32}P]tRNA was used as probe for tRNA genes. The 300-bp insert of *BamHI*-digested BLUR8 plasmid DNA (Rubin *et al.*, 1980) was separated from the pBR322 vector DNA on an agarose gel, nick translated (Rigby *et al.*, 1977), separated from the unincorporated [α-^{32}P]dCTP on a centrifugal gel filtration column (Mini-Spin from Worthington) and used as hybridization probe to detect any putative *Alu*-elements in the human DNA segment of λhVKV7. Southern hybridizations to nitrocellulose-immobilized DNA fragments were done in 50% formamide, 5x SSC (1x SSC is 150 mM NaCl and 15 mM sodium citrate, pH 7.0), 1 mM EDTA, 0.1% SDS, 2 μg/ml carrier tRNA and [3'-^{32}P]tRNA at 43°C. Prehybridization of rinsed and air-dried alkaline transfers to charged-modified nylon membranes were done overnight before adding ^{32}P-labeled probe to the prehybridization solution. For DNA-RNA hybridizations with nylon membrane-immobilized DNA, 50% formamide, 1.5X SSPE (1X SSPE is 10 mM sodium phosphate, 180 mM NaCl, 1 mM EDTA, pH 7.4), 1% SDS, 0.5% lowfat dry milk (Carnation), 2 mg/ml yeast carrier tRNA was used at 43°C or 50°C. For DNA-DNA hybridizations, 1.5X SSPE, 1% SDS, 0.5% lowfat dry milk and 10 mg/ml sheared salmon sperm carrier DNA was used at 68°C. After

prehybridization, the ^{32}P-labeled probe was added to the prehybridization solution, and incubation of the membrane continued 1-2 days. After hybridization, the filters were washed successively at room temperature for 15 min in the following solutions: (1) 2x SSC, 0.1% SDS; (2) 0.5x SSC, 0.1% SDS; and (3) 0.1x SSC, 0.1% SDS. The filters were air-dried and autoradiographed.

Nucleotide Sequence Analysis

The DNA sequence of a 7,078-bp region of the human DNA segment in λhVKV7 was determined by the dideoxynucleotide chain-termination procedure (Sanger *et al.*, 1977; Sanger *et al.*, 1980), and both DNA strands were sequenced to ensure accuracy. About one-half of this nucleotide sequence was done by Mike M. Lee. Either the Klenow fragment of *E. coli* DNA polymerase I (New England Biolabs) or the modified T7 DNA polymerase (Sequenase from U.S. Biochemical) was utilized in the sequencing reactions for chain extension from the synthetic oligodeoxynucleotide universal primer (New England Biolabs), corresponding to a 17-bp segment of the M13 vector DNA sequence, with either [^{35}S]deoxyadenosine 5'-(α-thio)-triphosphate (New England Nuclear/Dupont) or [α-^{32}P]dATP (ICN). In certain cases, in order to overlap sequences between some fragments, to sequence difficult areas in which no partial fragments could be isolated,

or to extend the sequence of the upstream region flanking the valine $tRNA_{CAC}$ gene, synthetic oligodeoxynucleotide primers were used instead of the universal sequencing primer. These primers (obtained from Philip Buzby and Maurice Kashdan of New England Nuclear/Dupont) were synthesized on a Dupont Coder 300 Synthesizer and used without further purification. Ambiguities due to compression effects were resolved by utilizing the dGTP analog dITP (Barnes *et al.*, 1983). Labeled fragments produced in the sequencing reactions were separated electrophoretically on 8%, 6%, or 4% denaturing, wedge-shaped (0.4 mm to 0.8 mm) polyacrylamide gels. The gels were dried and autoradiographed using Kodak XAR-5 film. Four contiguous DNA subfragments were isolated from λhVKV7 for DNA sequence analysis (depicted in Fig. 1): a 1.5-kb *SmaI* fragment (left side), a 1.6-kb *SmaI/BamHI* fragment (middle left), a 2.1-kb *BamHI* fragment (middle right),and a 1.5-kb *BamHI* fragment (right side). These four DNA fragments were subcloned into the corresponding restriction sites of M13mp18RF and/or M18mp19RF, and transformed into competent *E. coli* JM101 (Yanisch-Perron *et al.*, 1985). A series of overlapping deletion subclones for DNA sequence analysis was generated for the 1.5-kb *BamHI*, 2.1-kb *BamHI*, and 1.5-kb *SmaI* subfragments by the 3'-exonuclease activity of T4 DNA polymerase (Dale *et al.*, 1985) and for the 1.6-kb

SmaI/BamHI subfragment by exonuclease III (Henikoff, 1984) as depicted in the sequencing strategy in Fig. 1. The synthetic primers 5'-CGTCAGCCAGCTAGCTGG-3' (around residue 6658 in Fig. 2) and 5'-GCGAGACAAAGGCGGAAGC-3' (around residue 1520 in Fig. 2) were used to complete the nucleotide sequences of the 1.5-kb *BamHI* and 1.5-kb *SmaI* fragments, respectively. The overlapping fragment for sequencing across the junction between the 1.5-kb and 2.1-kb *BamHI* fragments was generated by digesting the 8.1-kb *EcoRI* fragment (Fig. 1A, right side) containing both *BamHI* fragments with *SmaI* and *SphI*. The resulting fragments were ligated into *SmaI/SphI*-digested M13mp19RF, and the subcloned DNAs were transformed into *E. coli* JM101. The progeny M13 phage DNAs were size selected for the 1.4-kb *SmaI/SphI* overlapping fragment, which was sequenced from the *SmaI* end. The overlapping fragments for sequencing across the junctions between the 2.1-kb *BamHI* and 1.6-kb *SmaI/BamHI* fragments and between the 1.6-kb *SmaI/BamHI* and 1.5-kb *SmaI* fragments were generated by digesting the 5.6-kb *AccI* fragment (Fig. 1A, right middle portion) with *HindIII* and *KpnI*. The resulting fragments were ligated into *HindIII/KpnI*-digested M13mp18RF and M13mp19RF, and the DNAs were transformed into *E. coli* JM101. The M13 phage DNAs were screened for the overlapping 1.3-kb *KpnI/HindIII* and 1.9-kb *HindIII/KpnI* fragments (Fig. 1A, middle left section). The M13mp18 subclone containing the 1.3-kb insert

was sequenced using the universal primer to confirm the overlap between the 2.1-kb *BamHI* and 1.6-kb *SmaI/BamHI* fragments. The M13mp19 subclone containing the 1.9-kb insert was sequenced using the primer 5'-CCAGGCTGAAGTGCAGTGG-3' (around residue 1909 in Fig. 2) to confirm the overlap between the 1.6-kb *SmaI/BamHI* and 1.5-kb *SmaI* fragments. After the completion of the sequence of the 1.5-kb *SmaI* fragment, it was necessary to extend the sequence, since only about 90 bp of the flanking region upstream from the valine $tRNA_{CAC}$ gene was contained in the 1.5-kb *SmaI* fragment. To accomplish this, the 5.6-kb *AccI* fragment (Fig. 1A) was filled-in with the Klenow fragment of *E. coli* DNA polymerase I (Maniatis *et al.*, 1982). One aliquot of the blunt-ended *AccI* fragment was digested with *KpnI*, and the resulting fragments cloned into *KpnI/SmaI*-digested M13mp19RF for transformation into *E. coli* JM101. The single-strand phage DNAs were size selected for the 2.0-kb *KpnI/AccI* insert. The primer 5'-GCCGGGATAGCTCTAGGC-3' (around nucleotide 473 in Fig. 2) was used to extend the sequence of one strand. To sequence the complementary strand, the other aliquot of the blunt-ended *AccI* fragment was cloned into the *SmaI* site of pUC18 for double-stranded sequencing (Chen and Seeburg, 1985; Zhang *et al.*, 1988). The amplified and HPLC-purified plasmid DNA was passed through an Elutip-d column (Schleicher and Schuell) prior to

the sequencing reactions. The primers 5'-CATCTACATTCTCACAGC-3' and 5'-CTCAAGGCGGCTCGCGGG-3' (around residues 1 and 237, respectively in Fig. 2) were utilized to finish the DNA sequence.

The DNA sequence was stored and analyzed by Irma Pirtle with microcomputer programs (Larson and Messing, 1983; Pearson and Lipman, 1988). The three tRNA genes were identified with programs that search for tRNA genes in long nucleotide sequences (Shortridge *et al.*, 1986).

In Vitro Transcription Analysis

To assay for the transcriptional efficiencies of the tRNA genes, three subclones containing the individual tRNA genes and excluding the *Alu* elements were constructed and purified by Susan K. Legan by ligating specific restriction fragments into plasmid DNA. A subclone with a 641-bp *RsaI* fragment containing the valine $tRNA_{CAC}$ gene inserted into the *SmaI* site of pUC18, a subclone with a 289-bp *RsaI* fragment containing the lysine $tRNA_{CUU}$ gene inserted into the *SmaI* site of pUC19, and a subclone with a 1,083-bp *KpnI/BamHI* fragment encompassing the valine $tRNA_{AAC}$ gene inserted into pUC19, were designated p18V7C, p19K7R, and p19V7A, respectively. In addition, a subclone with a 746-bp *SmaI/EcoRI* fragment containing both the lysine $tRNA_{CUU}$ gene and the valine $tRNA_{CAC}$ gene was ligated into pUC19 and

designated p19KV7ES. These hybrid molecules were used to transform competent *E. coli* DH5α (Hanahan, 1983), and grown in TB media (89 mM potassium phosphate pH 7.5, 12 g/l tryptone, 24 g/l yeast extract, 0.4% glycerol (v/v), and 25 µg/l ampicillin). Crude plasmid DNA was prepared (Pirtle *et al.*, 1986) and purified on a Waters DEAE-5PW anion-exchange HPLC column (7.5 mm x 7.5 cm) using 20 mM potassium phosphate pH 6.9, 5 M urea, and a linear gradient of KCl (0.3-0.6 M) for 50 min at 0.8 ml/min. The junctions between the plasmid vector and human inserts were sequenced to confirm the end points (Zhang *et al.*, 1988).

The plasmid DNAs were transcribed by Irma Pirtle and Robert Pirtle in 25-µl assay mixtures containing 2.5-20 nM plasmid DNA, sufficient non-specific pUC18 or pUC19 DNA to adjust the final DNA concentration to 80 µg/ml, 10 µCi of [α-^{32}P]GTP (650 Ci/mmole), 20 mM HEPES pH 7.9, 6 mM creatine phosphate, 0.14 mM EDTA, 5.6 mM $MgCl_2$, 0.8 mM DTT, 20 mM ammonium sulfate, 68 mM KCl,α-amanitin (2 µg/ml), 60 µM each of CTP, UTP, and ATP, 20 µM GTP, 10% glycerol (v/v), 5 µl of HeLa cell lysate from BRL (9 mg protein/ml) prepared by the cytoplasmic extraction procedure (Weil *et al.*, 1979), and 10 µl of HeLa cell lysate from BRL (24 mg protein/ml) prepared by the whole-cell extraction procedure (Manley *et al.*, 1980). After incubation at 30°C for 90 min, 175 µl of 0.2% SDS, 0.3 M sodium acetate pH 5.2, and 20 µg/ml yeast carrier

tRNA was added. The mixtures were extracted with phenol/chloroform (1:1), precipitated, and redissolved in 50% formamide gel loading dye for electrophoresis on 6% denaturing polyacrylamide gels. Gel slices with the ^{32}P-labeled tRNA transcription products were excised and counted by Cerenkov radiation to quantitate the relative amounts of RNA. Mature-sized tRNA transcripts were digested with RNase T_1 and fingerprinted by two-dimensional electrophoresis-homochromatography (Pirtle *et al.*, 1980).

Results

Characterization of the tRNA Genes and *Alu* Elements in λhVKV7

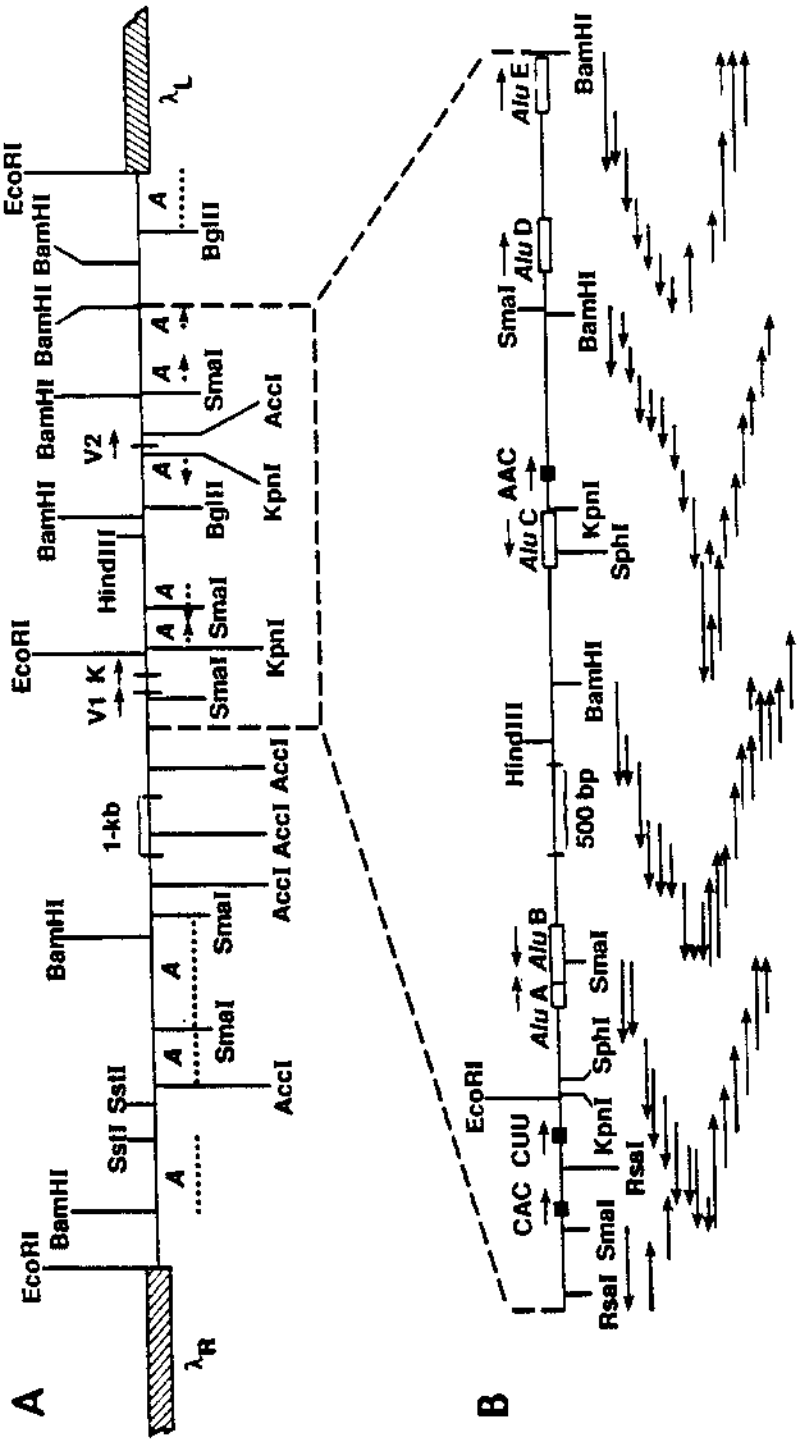
The lambda clone λhVKV7 from a human genomic library was found to have at least two presumptive tRNA genes, because two hybridizing bands were observed in Southern blots of *SmaI* (3.5-kb and 1.5-kb), *BamHI* (7.1-kb and 2.1-kb), and *EcoRI* (8.1-kb and 10.4-kb) digestions of λhVKV7 DNA (Wang, 1989). These same bands also hybridized to the purified valine[3'-^{32}P]tRNA$_{TAC}$, tentatively indicating that the clone harbored two or more valine tRNA genes. The physical map derived for the 18.5-kb segment of human DNA is shown in Fig. 1A.

The human DNA segment of λhVKV7 has nine or more *Alu* family members, as indicated from hybridization of ^{32}P-labeled nick-translated DNA fragments derived from the 300-bp insert of BLUR8 (Rubin *et al.*, 1980) to alkaline transfers of restricted λhVKV7 DNA (Wang, 1989) and also by DNA sequence analysis.

Nucleotide sequence encompassing three tRNA genes and five *Alu* elements

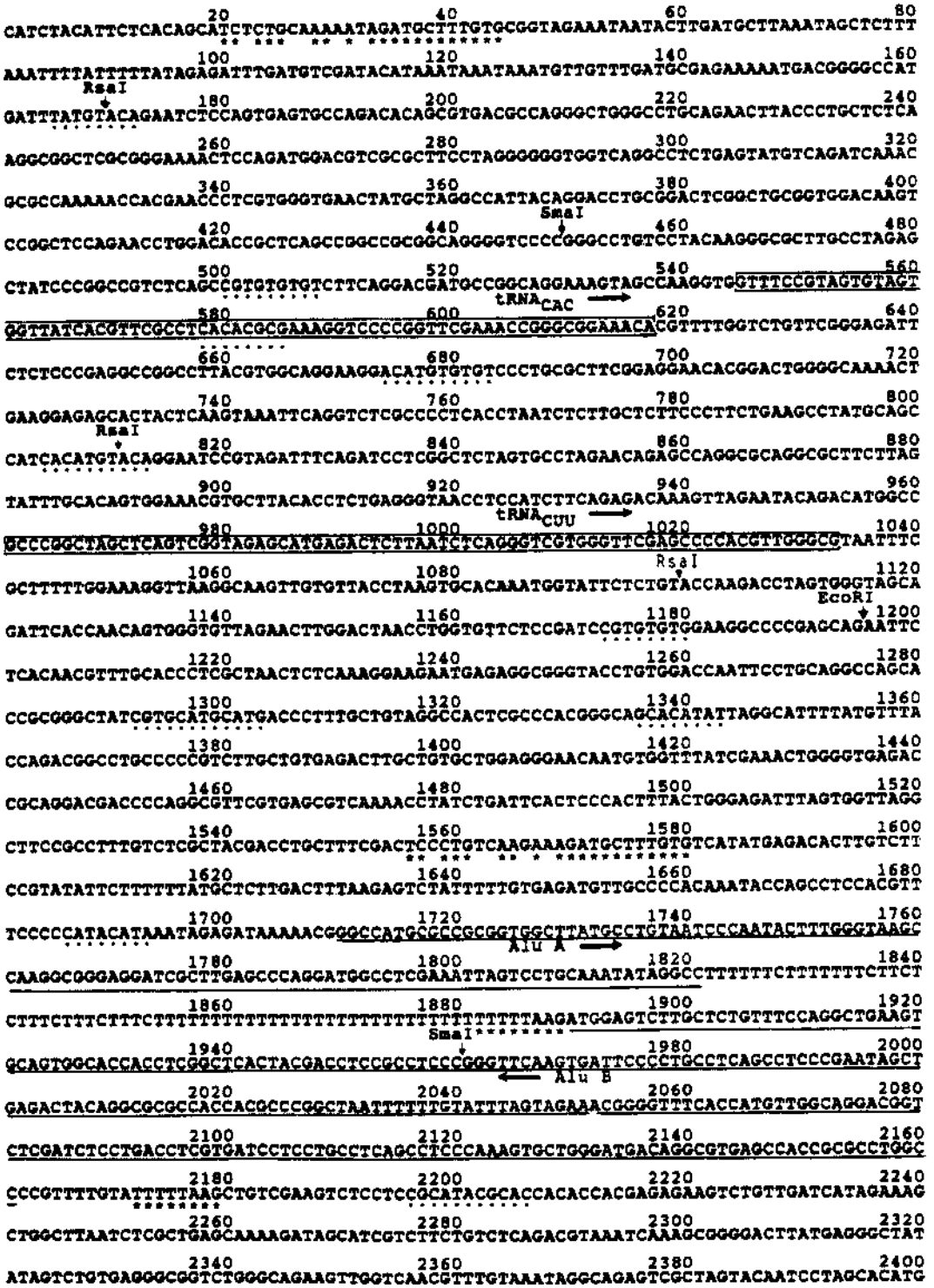
The strategy used to sequence the 7078-bp region of

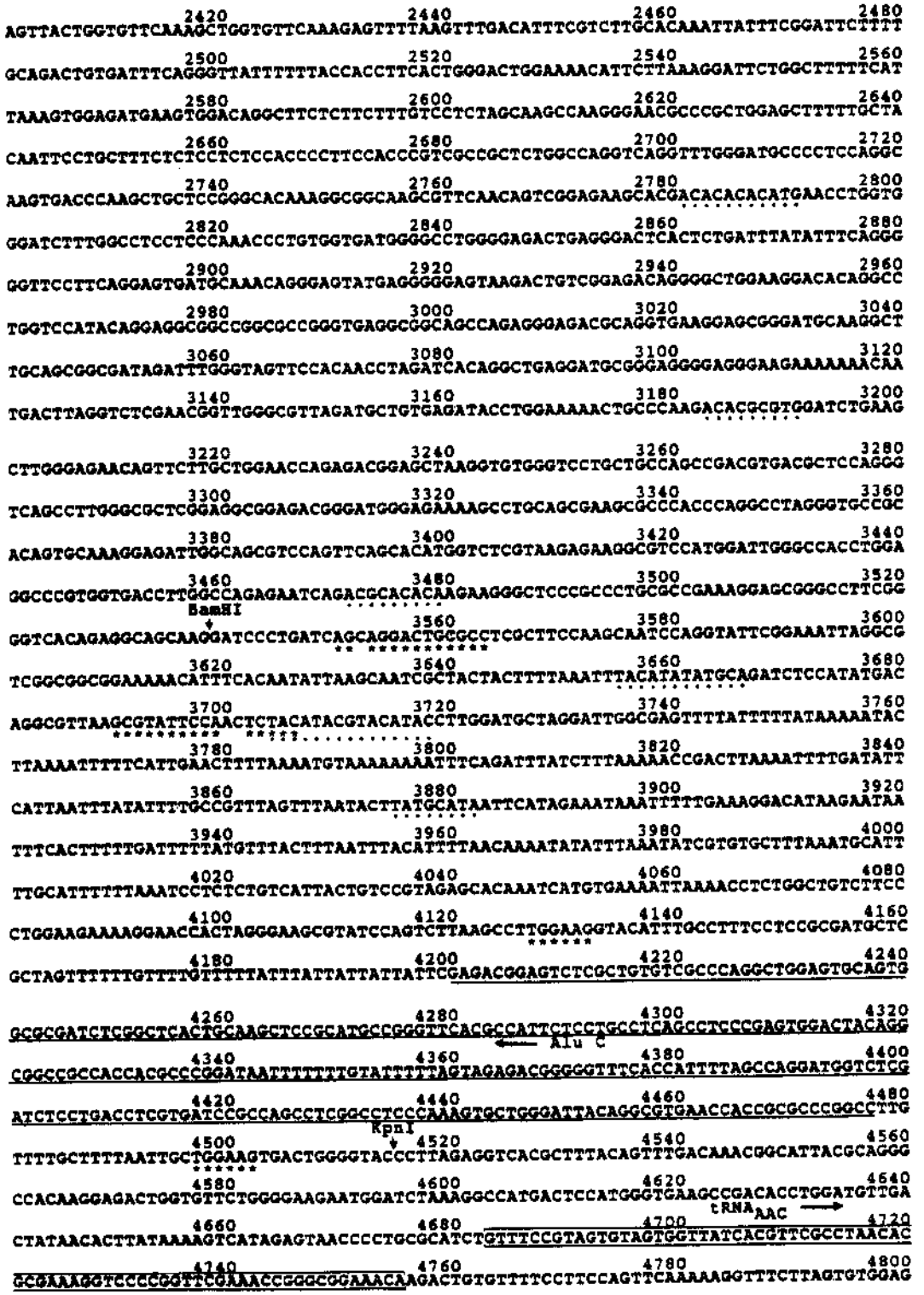
Fig. 1. Physical map of the human genomic clone λhVKV7 and DNA sequencing strategy. **A**. Physical map of the 18.5-kb human DNA segment harbored in λhVKV7, encompassing two valine tRNA genes, a lysine tRNA gene, and at least nine *Alu* elements. The horizontal line represents human DNA and the hatched areas represent the λ Charon-4A vector arms. The black rectangles represent the tRNA structural genes, the valine $tRNA_{CAC}$ gene indicated by V1, the lysine $tRNA_{CUU}$ gene by K, and the valine $tRNA_{AAC}$ gene by V2. The dotted lines represent the loci of the *Alu* elements. The arrows specify the direction of transcription of the tRNA genes and the relative polarities of the *Alu* members (from 5' to 3'). The locus of a 200-bp *AccI* fragment in the physical map was not determined. **B**. Sequencing strategy. The dashed lines indicate the 7,078-bp sequenced region of λhVKV7. The black rectangles represent the tRNA genes, with the respective anticodons indicated above the genes. The open rectangles represent the loci of *Alu* elements. The arrows at the bottom indicate the direction and extent of the deletion subfragments (Henikoff, 1984; Dale *et al.*, 1985) sequenced or of the synthetic oligodeoxynucleotide primer-extended sequenced regions. The DNA sequence was determined by sequencing four contiguous fragments and by extending the sequence of the 1.5-kb *SmaI* fragment. The overlaps between the four fragments were sequenced on one strand only.


A
B
EcoRI
BamHI
SstI
SstII
AccI
SmaI
1-kb
V1 K
HindIII
V2
KpnI
BglII
λ_R
λ_L
CAC CUU
RsaI
Alu A
Alu B
Alu C
Alu D
Alu E
AAC
SphI
500 bp

λhVKV7 shown in Fig. 2 is depicted in Fig. 1B. Four adjoining DNA subfragments were selected for DNA sequence analysis, since two of the subfragments hybridized to unfractionated mammalian tRNA probe, and all four subfragments hybridized to an *Alu* probe. In addition, the sequence of several hundred base pairs beyond the *SmaI* site on the left of Fig. 1B was determined, to extend the upstream region flanking the valine $tRNA_{CAC}$ gene. Three presumptive tRNA genes (Fig. 2) were identified within the sequenced region of λhVKV7 by computer analysis (Shortridge *et al.*, 1986). A valine $tRNA_{AAC}$ gene (Figs. 2 and 3A) was found within the 2.1-kb *BamHI* fragment that hybridized to the mammalian valine $tRNA_{IAC}$ probe. A lysine $tRNA_{CUU}$ gene (Figs. 2 and 3B) is about 3.6 kb further upstream from the valine $tRNA_{AAC}$ gene, and a valine $tRNA_{CAC}$ gene (Figs. 2 and 3A) is 342 bp upstream from the lysine tRNA gene, such that both tRNA genes are harbored within the 1.5-kb *SmaI* fragment that hybridized to the valine $tRNA_{IAC}$ probe (Fig. 1B). The tRNA genes all have the same relative polarity in the human DNA segment. The nucleotide sequences of the noncoding strands of the valine tRNA genes are depicted in the cloverleaf conformation in Fig. 3A, and that of the lysine $tRNA_{CUU}$ gene is shown in Fig. 3B. The three tRNA genes have all the invariant and semi-invariant nucleotides found in

Fig. 2. Nucleotide sequence of a region of the human DNA segment of λhVKV7 encompassing a cluster of tRNA genes and *Alu* elements. The sequences of the coding regions of the valine $tRNA_{CAC}$, lysine $tRNA_{CUU}$, and valine $tRNA_{AAC}$ genes are boxed. The sequences of the *Alu* elements are underlined. The arrows indicate the direction of transcription of the three tRNA genes and the relative polarities of *Alu* members A to E (from 5' to 3'). The tracts of eight or more nucleotides with the typical alternating purine-pyrimidine motif which have the potential to adopt the Z-DNA conformation are indicated by the dotted underlines. Direct repeat sequences are indicated by asterisks.


RsaI
SmaI
tRNA CAC
tRNA CUU
RsaI
EcoRI
Alu A
SmaI
Alu B


BamHI
KpnI
Alu C
tRNA AAC

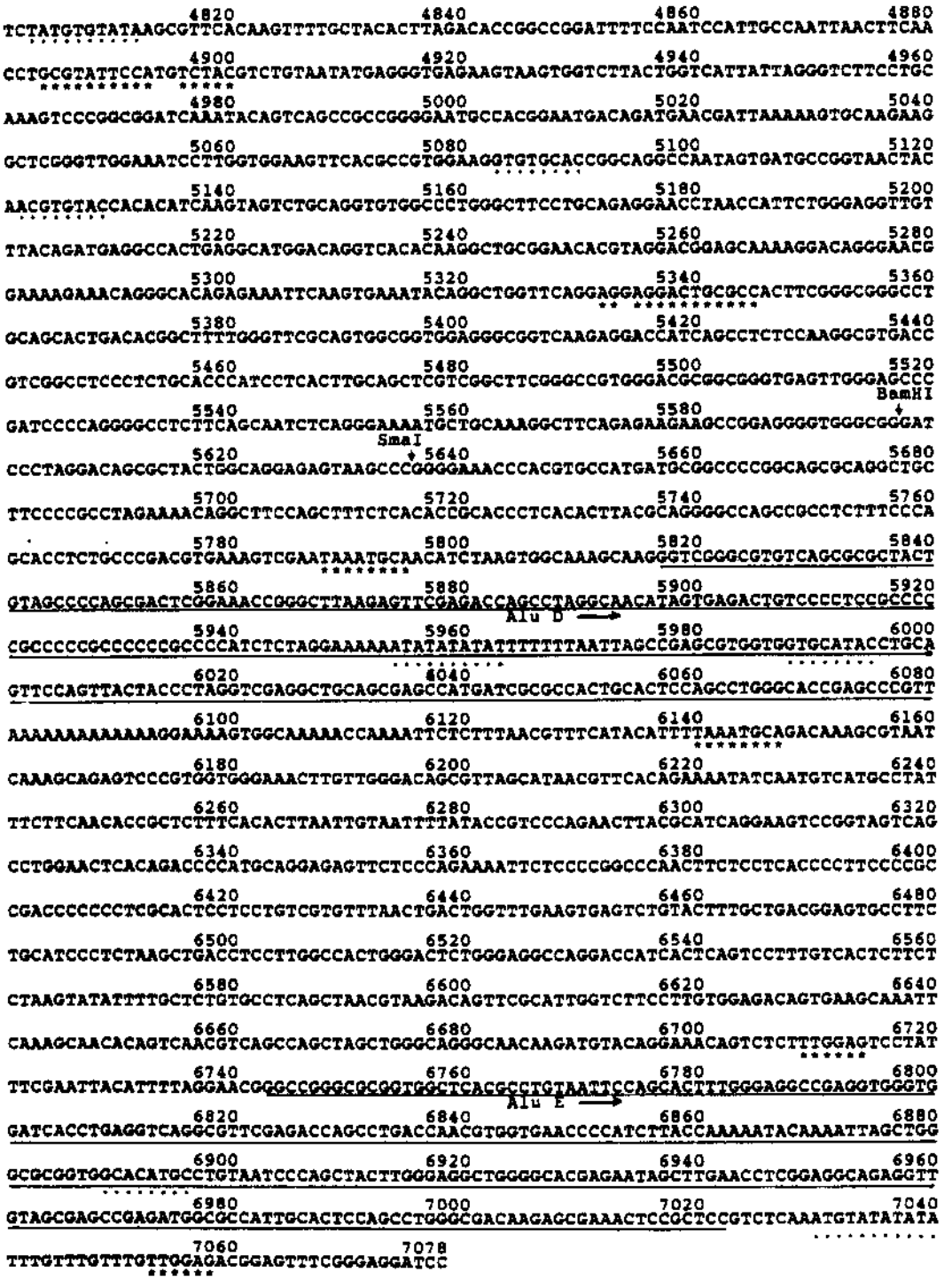
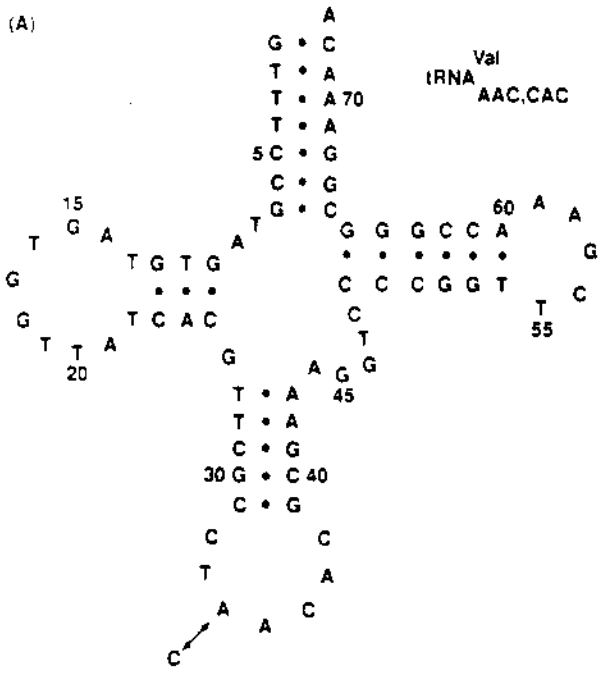
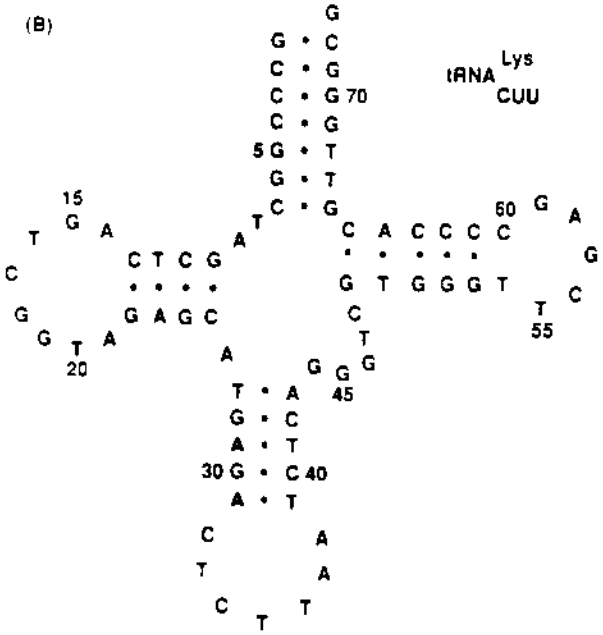

4820 4840 4860 4880
TCTATGTGTATAAGCGTTCACAAGTTTTGCTACACTTAGACACCGGCCGGATTTTCCAATCCATTGCCAATTAACTTCAA
4900 4920 4940 4960
CCTGCGTATTCCATGTCTACGTCTGTAATATGAGGGTGAGAAGTAAGTGGTCTTACTGGTCATTATTAGGGTCTTCCTGC
4980 5000 5020 5040
AAAGTCCCGGCGGATCAAATACAGTCAGCCGCCGGGGAATGCCACGGAATGACAGATGAACGATTAAAAAGTGCAAGAAG
5060 5080 5100 5120
GCTCGGGTTGGAAATCCTTGGTGGAAGTTCACGCCGTGGAAGGTGTGCACCGGCAGGCCAATAGTGATGCCGGTAACTAC
5140 5160 5180 5200
AACGTGTACCACACATCAAGTAGTCTGCAGGTGTGGCCCTGGGCTTCCTGCAGAGGAACCTAACCATTCTGGGAGGTTGT
5220 5240 5260 5280
TTACAGATGAGGCCACTGAGGCATGGACAGGTCACACAAGGCTGCGGAACACGTAGGACGGAGCAAAAGGACAGGGAACG
5300 5320 5340 5360
GAAAAGAAACAGGGCACAGAGAAATTCAAGTGAAATACAGGCTGGTTCAGGAGGAGGACTGCGCCACTTCGGGCGGGCCT
5380 5400 5420 5440
GCAGCACTGACACGGCTTTTGGGTTCGCAGTGGCGGTGGAGGGCGGTCAAGAGGACCATCAGCCTCTCCAAGGCGTGACC
5460 5480 5500 5520
GTCGGCCTCCCTCTGCACCCATCCTCACTTGCAGCTCGTCGGCTTCGGGCCGTGGGACGCGGCGGGTGAGTTGGGAGCCC
BamHI
5540 5560 5580
GATCCCCAGGGGCCTCTTCAGCAATCTCAGGGAAAATGCTGCAAAGGCTTCAGAGAAGAAGCCGGAGGGGTGGGCGGGAT
SmaI
5620 5640 5660 5680
CCCTAGGACAGCGCTACTGGCAGGAGAGTAAGCCCGGGGAAACCCACGTGCCATGATGCGGCCCCGGCAGCGCAGGCTGC
5700 5720 5740 5760
TTCCCCGCCTAGAAAACAGGCTTCCAGCTTTCTCACACCGCACCCTCACACTTACGCAGGGGCCAGCCGCCTCTTTCCCA
5780 5800 5820 5840
GCACCTCTGCCCGACGTGAAAGTCGAATAAATGCAACATCTAAGTGGCAAAGCAAGGGTCGGGCGTGTCAGCGCGCTACT
5860 5880 5900 5920
GTAGCCCCAGCGACTCGGAAACCGGGCTTAAGAGTTCGAGACCAGCCTAGGCAACATAGTGAGACTGTCCCCTCCGCCCC
Alu D
5940 5960 5980 6000
CGCCCCCGCCCCCCGCCCCATCTCTAGGAAAAAATATATATATTTTTTTAATTAGCCGAGCGTGGTGGTGCATACCTGCA
6020 6040 6060 6080
GTTCCAGTTACTACCCTAGGTCGAGGCTGCAGCGAGCCATGATCGCGCCACTGCACTCCAGCCTGGGCACCGAGCCCGTT
6100 6120 6140 6160
AAAAAAAAAAAAAAGGAAAAGTGGCAAAAACCAAAATTCTCTTTAACGTTTCATACATTTTAAATGCAGACAAAGCGTAAT
6180 6200 6220 6240
CAAAGCAGAGTCCCGTGGTGGGAAACTTGTTGGGACAGCGTTAGCATAACGTTCACAGAAAATATCAATGTCATGCCTAT
6260 6280 6300 6320
TTCTTCAACACCGCTCTTTCACACTTAATTGTAATTTTATACCGTCCCAGAACTTACGCATCAGGAAGTCCGGTAGTCAG
6340 6360 6380 6400
CCTGGAACTCACAGACCCCATGCAGGAGAGTTCTCCCAGAAAATTCTCCCCGGCCCAACTTCTCCTCACCCCTTCCCCGC
6420 6440 6460 6480
CGACCCCCCCTCGCACTCCTCCTGTCGTGTTTAACTGACTGGTTTGAAGTGAGTCTGTACTTTGCTGACGGAGTGCCTTC
6500 6520 6540 6560
TGCATCCCTCTAAGCTGACCTCCTTGGCCACTGGGACTCTGGGAGGCCAGGACCATCACTCAGTCCTTTGTCACTCTTCT
6580 6600 6620 6640
CTAAGTATATTTTGCTCTGTGCCTCAGCTAACGTAAGACAGTTCGCATTGGTCTTCCTTGTGGAGACAGTGAAGCAAATT
6660 6680 6700 6720
CAAAGCAACACAGTCAACGTCAGCCAGCTAGCTGGGCAGGGCAACAAGATGTACAGGAAACAGTCTCTTTGGAGTCCTAT
6740 6760 6780 6800
TTCGAATTACATTTTAGGAACGGGCCGGGCGCGGTGGCTCACGCCTGTAATTCCAGCACTTTGGGAGGCCGAGGTGGGTG
Alu E
6820 6840 6860 6880
GATCACCTGAGGTCAGGCGTTCGAGACCAGCCTGACCAACGTGGTGAAACCCCATCTTACCAAAAATACAAAATTAGCTGG
6900 6920 6940 6960
GCGCGGTGGCACATGCCTGTAATCCCAGCTACTTGGGAGGCTGGGGCACGAGAATAGCTTGAACCTCGGAGGCAGAGGTT
6980 7000 7020 7040
GTAGCGAGCCGAGATGGCGCCATTGCACTCCAGCCTGGGCGACAAGAGCGAAACTCCGCTCCGTCTCAAATGTATATATA
7060 7078
TTTGTTTGTTTGTTGGAGACGGAGTTTCGGGAGGATCC

Fig. 3. Sequences of human tRNA genes from the noncoding strand of λhVKV7 in the cloverleaf conformation. **A**. The valine $tRNA_{AAC}$ and $tRNA_{CAC}$ genes. The arrow denotes the anticodon wobble base at position 34, which is the only nucleotide in the major cytoplasmic valine $tRNA_{AAC}$ isoacceptor which differs from that of the minor cytoplasmic valine $tRNA_{CAC}$ isoacceptor. As noted by Jank *et al.* (1977b), tRNAs encoded by valine tRNA genes could potentially have a T54:A60 base pair in the T stem. **B**. The lysine $tRNA_{CUU}$ gene.


(A)
tRNA Val
AAC,CAC


(B)
tRNA Lys
CUU

mature tRNAs (Rich and RajBhandary, 1976), and have characteristic split internal promoter elements and also transcription termination sites of at least four consecutive T residues in the 3'-flanking regions (Geiduschek and Tocchini-Valentini, 1988). The tRNA genes do not encode the CCA trinucleotide found at the 3' end of mature tRNAs.

The sequenced regions flanking the tRNA genes in λhVKV7 have several notable strucural features. There are 23 tracts of from 8 to 15 nucleotides which have the potential to adopt the Z-DNA conformation (depicted by the dotted underlines in Fig. 2), since they have the typical alternating purine-pyrimidine motif (Rich *et al.*, 1984), and several occur within 200-300 bp of the tRNA genes. The flanking regions of the valine $tRNA_{CAC}$ gene are GC-rich, since the upstream region from -144 to -44 (residues 400-500 in Fig. 2) and the downstream region from +15 to +40 (residues 632-657) are both 73% GC. This is in contrast to the high AT-rich flanking regions of the two valine tRNA genes and valine tRNA pseudogene characterized by Arnold *et al.*(1986) and the valine tRNA gene variant described by Kahnt *et al.*(1989).

Four full-length *Alu* family members, arbitrarily designated *Alu* B-E, are interspersed with the three tRNA genes throughout the sequenced region of λhVKV7 (shown schematically in Fig. 1B and underlined in Fig. 2). The

flanking direct repeat sequences characteristic of target site reduplications upon retroposition (Weiner *et al.*, 1986) are denoted by asterisks in Fig. 2. In addition, a truncated *Alu* element (designated *Alu* A) that corresponds basically to the left monomer unit of the dimeric retropseudogene structural element was identified immediately downstream from the *Alu* B structural region. No target site direct repeats could be identified for *Alu* A due to the nearby retroposition of *Alu* B. A 67-bp homopyrimidine-homopurine stretch (nucleotides 1820-1886 in Fig. 2) occurs between *Alu* members A and B. Segments of DNA with such pyrimidine-purine strand bias are thought to adopt unusual non-B, right-handed DNA conformations (Wells, 1988). *Alu* members, A, D, and E are of the same relative polarity as the tRNA genes, while *Alu* members B and C are of opposite polarity within the human DNA segment. The homologies of *Alu* members B to E with the *Alu* consensus sequence derived by Quentin (1988) are 88%, 91%, 46%, and 88%, respectively, while the truncated *Alu* member A has a homology of 60% with the *Alu* left monomer unit.

In Vitro Transcription Analysis

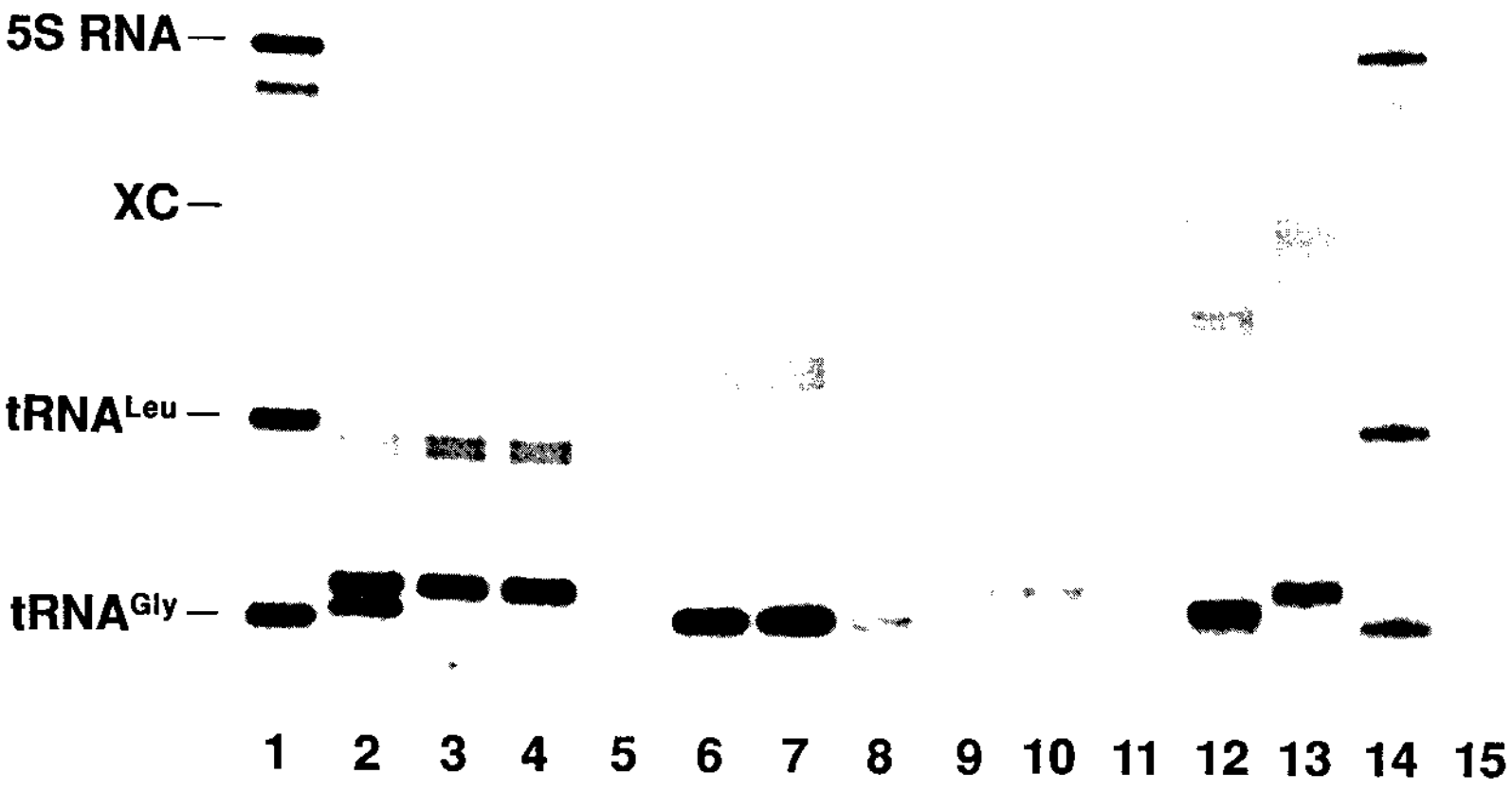
The three individually subcloned tRNA genes and the plasmid subclone harboring both the valine $tRNA_{CAC}$ gene and the lysine $tRNA_{CUU}$ gene were transcribed in a homologous *in*

vitro HeLa cell-free system, and the transcription products were resolved on 6% denaturing polyacrylamide gels, an example of which is shown in Fig. 4. All three tRNA genes successfully direct transcription and generate products with electrophoretic mobilities for precursor- and mature-sized tRNAs. The estimated sizes of the mature-sized valine $tRNA_{CAC}$ transcription product (lanes 3 and 4), the lysine $tRNA_{CUU}$ transcription product (lanes 6 and 7), and the valine $tRNA_{AAC}$ transcription product (lanes 9 and 10) are all about 75-76 nucleotides, in comparison to the sizes of mammalian glycine [3'-^{32}P]tRNA (75 nucleotides) and leucine [3'-^{32}P]tRNA (86 residues) standards (lanes 1 and 14), and also to the mature-sized tRNA transcription products corresponding to a glycine tRNA (lane 12, 74 residues) and a valine $tRNA_{AAC}$ (lane 13, 76 residues) characterized previously (Pirtle *et al.*, 1986; Shortridge *et al.*, 1989). Presumptive precursor tRNA transcripts migrate with slightly lower mobilities than the mature-sized tRNA transcripts in Fig. 4, having chain lengths in the size range of 85-92 nucleotides.

Intriguingly, the valine $tRNA_{CAC}$ gene apparently has a higher transcriptional efficiency than the valine $tRNA_{AAC}$ gene, since the valine $tRNA_{CAC}$ products were eightfold greater than the valine $tRNA_{AAC}$ products. The lysine tRNA

Fig. 4. Fractionation of *in vitro* transcription products on a 6% denaturing polyacrylamide gel. Plasmid DNAs containing the subcloned individual tRNA genes were transcribed for 90 min at 30°C in a homologous Hela cell *in vitro* transcription system. Lanes 1 and 14 contain as standards [3'-^{32}P]-labeled *E. coli* 5S rRNA (121 nucleotides), bovine liver leucine $tRNA_{IAG}$ (86 nucleotides), and bovine liver glycine $tRNA_{GCC}$ (75 nucleotides). The transcription products are shown derived from the valine $tRNA_{CAC}$ gene of λhVKV7 harbored in the plasmid designated p18V7C at a concentration of 2.5 nM (lane 3), 5 nM (lane 4), and 5nM in the presence of 100 μg/ml α-amanitin (lane 5). The transcription products are also shown derived from the lysine $tRNA_{CUU}$ gene harbored in the plasmid designated p19K7R at a concentration of 2.5 nM (lane 6), 5 nM (lane 7), and 5 nM in the presence of 100 μg/ml α-amanitin (lane 8). The transcription products are shown derived from the valine $tRNA_{AAC}$ gene harbored in the plasmid designated p19V7A at a concentration of 2.5 nM (lane 9), 5 nM (lane 10), and 5 nM in the presence of 100 μg/ml α-amanitin (lane 11). Lane 2, transcription products derived from the plasmid designated p19KV7ES, encompassing both the valine $tRNA_{CAC}$ and lysine $tRNA_{CUU}$ genes. Lanes 12 and 13, products derived from transcription of plasmid DNAs at 5 nM harboring previously

characterized human tRNA genes (Pirtle *et al.*, 1986; Shortridge *et al.*, 1989) encoding a glycine $tRNA_{GCC}$ (lane 12) and a valine $tRNA_{AAC}$ (lane 13). No spurious products were derived from transcription of the plasmid pUC18 at 80 µg/ml (lane 15).


5S RNA
XC
tRNA^Leu
tRNA^Gly
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15

gene was transcribed slightly more efficiently than the valine $tRNA_{CAC}$ gene, since the lysine tRNA products were 1.4 times higher than the valine $tRNA_{CAC}$ products. In contrast, the valine $tRNA_{CAC}$ gene was transcribed about 1.4 times more efficiently than the lysine tRNA gene when the plasmid subclone harboring both genes was transcribed (lane 2). In the case in which the valine $tRNA_{CAC}$ and lysine tRNA genes were transcribed from the same plasmid DNA, the results could potentially be misleading since the competition between the two genes for RNA polymerase III and ancilliary trancription factors could potentially reduce the transcriptional efficiencies. Furthermore, the close proximity and relative polarities of the two genes could have effects on the observed transcriptional efficiencies (Russo *et al.*, 1987). The valine $tRNA_{CAC}$ and lysine tRNA genes are apparently transcribed (although to a much lesser extent) in the presence of 100 μg/ml of α-amanitin (lanes 5 and 8, respectively), a concentration typically used for inhibition of RNA polymerase III, which indeed does eliminate any detectable transcription of the valine $tRNA_{AAC}$ gene (lane 11). It is tempting to speculate that a *cis*-acting modulatory element occurs in the 5'-flanking region of the valine $tRNA_{CAC}$ gene that could be implicated in the apparently high transcriptional efficiency of the gene

relative to the valine $tRNA_{AAC}$ gene.

The two-dimensional RNase T_1 fingerprints (Fig. 5) of the three mature-sized tRNA transcription products are consistent with the predicted oligonucleotides deduced from the tRNA structural genes (Fig. 3), and also with the expected degradation products from the two mammalian valine tRNAs (Piper, 1975; Chen and Roe, 1977; Jank *et al.*, 1977a) and the mammalian lysine $tRNA_{CUC}$ (Raba *et al.*, 1979; Hedgcoth *et al.*, 1984). The two valine tRNA transcription products have essentially identical RNase T_1 fingerprints (Figs. 5A, B), with the exception of spot 13, which corresponds to the nonanucleotide derived from the anticodon region of the valine $tRNA_{CAC}$ (Fig. 5B). The valine $tRNA_{AAC}$ RNase T_1 fingerprint (Fig. 5A) has no similar spot, but has a pentanucleotide (5'-ACACG-3', spot 7), derived from the anticodon region instead of the nonanucleotide (5'-CCUAACACG-3'). The occurrence of this pentanucleotide in the RNase T_1 fingerprint of the valine $tRNA_{AAC}$ presumably could result from transglycosylation of the tRNA at A34 to I34 by the enzyme hypoxanthine insertase (Elliott and Trewyn, 1984), and cleavage at inosine by RNase T_1 to produce the unlabeled fragment 5'-CCUI-3' and the labeled 5'-ACACG-3' (visualized in spot 7). The RNase T_1 fingerprint of the valine $tRNA_{CAC}$ transcript has an identical pentanucleotide (spot 7 in

Fig. 5. RNase T_1 fingerprints of the mature-sized tRNA transcripts derived from the three tRNA genes in λhVKV7 by two-dimensional electrophoresis-homochromatography. The first dimension was electrophoresis on cellulose acetate at pH 3.5 for 45 min at 4,800 V. The second dimension was homochromatography at pH 7.5 and 65°C on DEAE-cellulose plates using homomixture C-30 (Pirtle *et al.*, 1980). The oligonucleotides were identified by relative mobility (Pirtle *et al.*, 1980) and by comparison to RNase T_1 fingerprints of the two mammalian valine tRNAs (Piper, 1975; Jank *et al.*, 1977a) and the lysine $tRNA_{CUU}$ (Raba *et al.*, 1979). The B with broken circle represents the location of the blue dye marker (xylene cyanol FF). **A.** Fractionation of the RNase T_1 digestion products derived from the valine $tRNA_{AAC}$ transcript. The oligonucleotides were assigned as: 1, Gp; 2, CpGp; 3, UpGp; 4, pGp; 5, UpApGp; 6, UpUpCpGp; 7, ApCpApCpGp; 8, ApApApGp: 9, UpUpUpCpCpGp; 10, m^7GpUpCpCpCpCpGp; 11, ApApApCpCpGp; and 12, UpUpApUpCpApCpGp. **B.** Fractionation of RNase T_1 hydrolysis products derived from the valine $tRNA_{CAC}$ transcript. Oligonucletides in spots 1-12 are identical to those in A, with the exception of spot 13, which was identified as CpCpUpCpApCpApCpGp. The occurrence of m^7G in the oligonucleotide designated as spot 10 in A and B was presumed to acount for the unusual mobility of the

oligonucleotide. Such base modification of m^7G has previously been observed in Hela cell extracts (Laski *et al.*, 1983). Other than this case, post-transcriptional modifications were discounted. **C**. Fractionation of the RNase T_1-derived oligonucletides from the lysine $tRNA_{CUU}$ transcript. The fragments were assigned as: 1, Gp; 2, CpGp; 3, UpGp; 4, pGp; 5, ApGp; 6, UpCpGp; 7, UpApGp; 8, UpUpGp; 9, UpUpCpGp; 10, CpCpCpGp; 11, the isomers CpUpApGp and CpApUpGp; 12, CpUpCpApGp; 13, CpCpCpCpApCpGp; and 14, ApCpUpCpUpUpApApUpCpUpCpApGp.

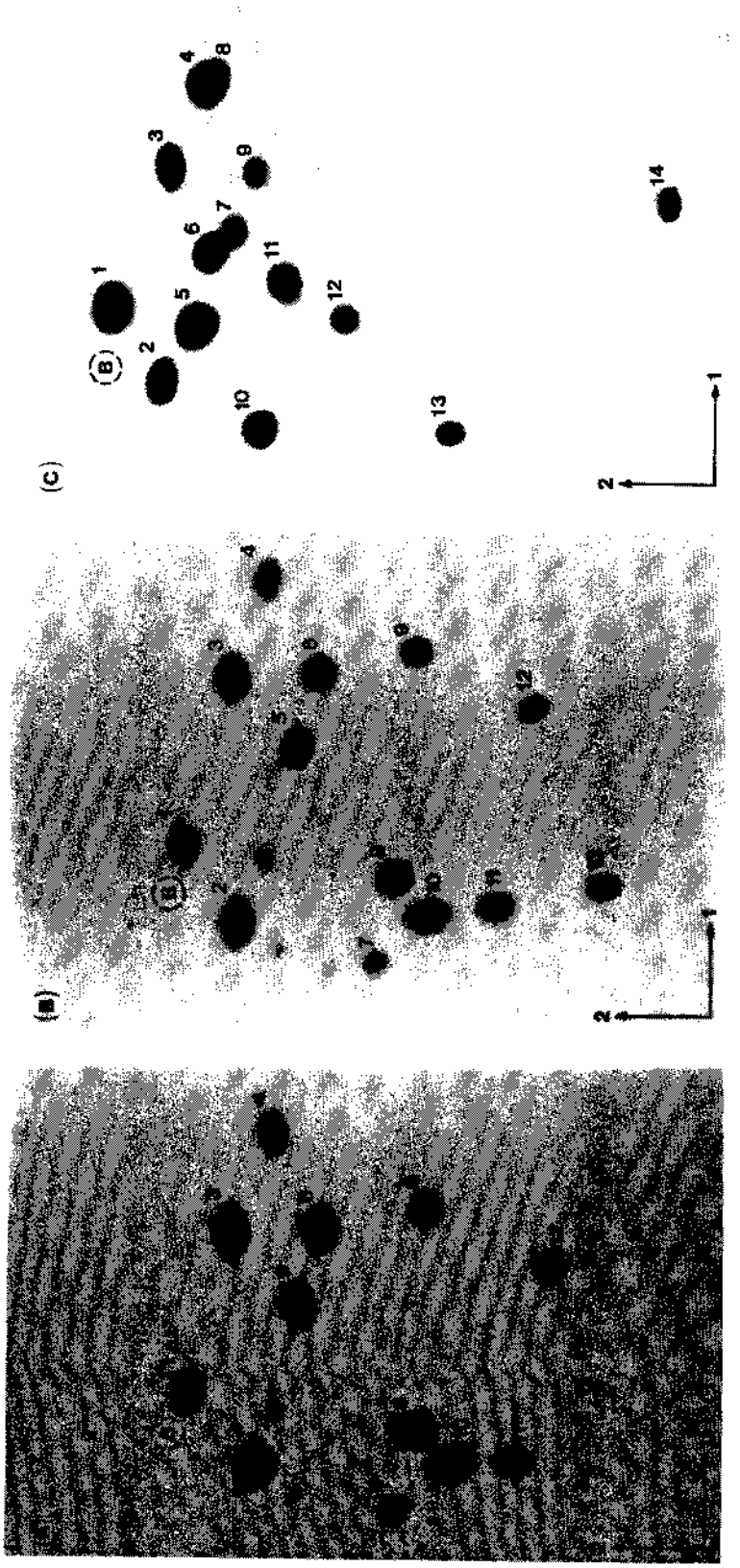

(C)
(B)
1
2
3
4
5
6
7
8
9
10
11
12
13
14
(B)
(A)

Fig. 5B). This could imply that a pyrimidine C34 to purine I34 transglycosylation occurred in a portion of the valine $tRNA_{CAC}$ transcript, such that both the pentamer and nonamer (spots 7 and 13, respectively) are seen in Fig. 5B. Since the exact substrate specificities of the hypoxanthine insertase have not been fully defined (Elliott and Trewyn, 1984), it is plausible that such a cytosine for hypoxanthine base exchange might occur in a portion of the minor valine $tRNA_{CAC}$ isoacceptor, rendering it into the major valine $tRNA_{IAC}$ isoacceptor.

Discussion

The tRNA gene cluster in λhVKV7 is unrelated to other human tRNA gene clusters characterized (Roy *et al.*, 1982; Chang *et al.*, 1986; Doran *et al.*, 1987; Shortridge *et al.*, 1989), since the physical maps differ greatly and since the tRNA isoacceptors encoded are different in each cluster. The tRNA gene cluster in λhVKV7 apparently contains two of the 13 members of the dispersed valine tRNA gene family, since the sizes of the 8.1-kb and 10.4-kb *EcoRI* fragments from λhVKV7 that hybridize to mammalian valine $tRNA_{IAC}$ are consistent with the sizes of bands E and C, respectively, detected by Arnold *et al.* (1986) in genomic blots of *EcoRI*-digested human DNA with mammalian valine $tRNA_{IAC}$ probe. Thus, the human DNA segment in λhVKV7 is very likely a single-copy segment in the genome. The three tRNA genes in λhVKV7 appear to be *bona fide* genes, since the DNA sequences of the tRNA coding regions are identical to the RNA sequences of the cognate mammalian cytoplasmic tRNAs (Piper, 1975; Chen and Roe, 1977; Jank *et al.*, 1977a; Raba *et al.*, 1979; Hedgcoth *et al.*, 1984), and because the three tRNA genes accurately direct the synthesis of discrete tRNA species whose RNase T_1 fingerprints are consistent with the predicted oligonucleotides derived from the tRNA structural genes.

The two valine tRNA genes in λhVKV7 are identical in sequence to the respective solitary human valine $tRNA_{AAC}$ and valine $tRNA_{CAC}$ genes characterized by Arnold *et al.* (1986). The valine $tRNA_{AAC}$ gene differs by one nucleotide (T20 in the dihydrouridine loop instead of C20) from another human valine $tRNA_{AAC}$ gene (Shortridge *et al.*, 1989) and by four nucleotides (in the TΨC arm) from an apparent valine $tRNA_{AAC}$ pseudogene (Arnold *et al.*, 1986). The valine $tRNA_{CAC}$ gene differs by four nucleotides from a variant human valine $tRNA_{CAC}$ gene (Kahnt *et al.*, 1989) and by five nucleotides from a *Xenopus* valine $tRNA_{CAC}$ gene (Peterson, 1987). The lysine $tRNA_{CUU}$ gene in λhVKV7 would encode the mammalian lysine $tRNA_2$ (Raba *et al.*, 1979; Hedgcoth *et al.*, 1984) and a tRNA identical to *Drosophila* lysine $tRNA_2$ (Silverman *et al.*, 1979), discounting modifications. The lysine tRNA gene is identical in sequence to two rat lysine $tRNA_{CUU}$ genes (Sekiya *et al.*, 1982) and has a homology of 84% with three human lysine $tRNA_{UUU}$ genes (Roy *et al.*, 1982; Doran *et al.*, 1987). The lysine tRNA gene probably encodes the primer tRNA for the human endogenous retroviral genomes (HERV-K), since the 3'-terminal 16 nucleotides of the lysine tRNA gene (Fig. 3B) are complementary to the primer binding site in these proviral elements (Ono *et al.*, 1986). In

addition, it is interesting to note that the lysine $tRNA_{CUU}$ gene has a 68% identity with a human phenylalanine tRNA gene (Doran *et al.*, 1987), but a 74% identity to an *E. coli* phenylalanine tRNA (Barrell and Sanger, 1969). Since the sequences of the flanking regions of the five characterized valine tRNA genes and valine tRNA pseudogene vary greatly, they appear to represent distinct members of the human valine tRNA gene family. Likewise, very few apparent similarities occur between the flanking regions of the lysine tRNA genes.

An interesting question concerning tRNA gene expression in mammalian genomes is whether or not *Alu* elements and related SINEs, some of which are transcribed by RNA polymerase III (Elder *et al.*, 1981), in the vicinities of tRNA genes influence the efficiency of tRNA gene transcription. Rosen and Daniel (1988) reported that a B2 element near a rat phenylalanine tRNA gene was transcribed an order of magnitude more efficiently than the tRNA gene, since the SINE element forms a more stable transcription complex than the phenylalanine tRNA gene, competing with it for binding the RNA polymerase III transcription factors. Thus, it is possible that some *Alu* members have similar effects on transcription of nearby primate tRNA genes. However, *Alu* elements have been identified near only a few human tRNA genes (Arnold *et al.*, 1986; Doran *et al.*, 1987;

Shortridge *et al.*, 1989), possibly because only short sequences flanking the tRNA genes have generally been sequenced. *Alu* members were found to occur within several hundred base pairs of two solitary valine tRNA genes (Arnold *et al.*, 1986), whereas an *Alu* element was found to actually overlap another valine tRNA gene (Shortridge *et al.*, 1989). *Alu* member C is just 180 bp upstream and of opposite orientation to the valine $tRNA_{CAC}$ gene, and *Alu* elements A and B are downstream from the valine $tRNA_{CAC}$ and lysine tRNA genes in λhVKV7 (Fig. 1). This suggests the likelihood that *Alu* elements occur near many members of the human valine tRNA gene family.

The human DNA segment harbored in the genomic clone λhVKV7 apparently has been the locus of multiple retroposition events throughout its evolutionary history. At least nine *Alu* members were found to be interspersed throughout the 18.5-kb segment of human DNA in λhVKV7, when only four or five should occur, based on an average spacing of about 4 kb (Britten *et al.*, 1988). *Alu* elements are thought to have originated in the primate genome some 65 million years ago by integration and amplification of a progenitor 7SL RNA source sequence (Weiner *et al.*, 1986). The *Alu* source sequences are thought to have undergone divergence such that at least four major *Alu* classes or subfamilies have evolved in the human lineage during the

last 30 or 40 million years (Britten *et al.*, 1988; Jurka and Smith, 1988; Quentin, 1988). *Alu* members can be categorized into the four subfamilies from diagnostic nucleotides in the *Alu* consensus sequence, with the older subfamilies bearing the most homology to the progenitor 7SL sequence (Britten *et al.*, 1988; Jurka and Smith, 1988; Quentin, 1988). Thus, of the five sequenced *Alu* elements in λhVKV7, *Alu* members A and D can be assigned to the oldest *Alu* subfamily I (J), *Alu* members B and E to the intermediate subfamilies II (a) and III (c), respectively, and *Alu* member C to the most recent *Alu* subfamily IV (b), using the categories of Britten *et al.* (1988) and Jurka and Smith (1988). According to the model for orientation and site-specific *Alu* retroposition (Daniels and Deininger, 1985), the more recent *Alu* member B integrated near the central A-rich stretch between the two original monomer units comprising the more ancient *Alu* member A, such that only the left monomer unit of *Alu* A remained intact and that the 67-bp pyrimidine-rich segment (residues 1820-1886 in Fig. 2) was generated in some manner during the retroposition event. Since all four *Alu* subfamilies are represented in the human DNA fragment in λhVKV7, it can be speculated that this region of the genome has undergone substantial divergence during the evolution of the human lineage.

In addition to the target site direct repeats of the

Alu members, three other direct repeat sequences: (I) 5'-TC_CTG__AA_A_AGATGCTTTGTG-3', nucleotides 20-44 and 1556-1580; (II) 5'-AG_AGGACTGCGCC-3', nucleotides 3550-3563 and 5332-5345; and (III) 5'-GCGTATTCCA__TCTAC-3', nucleotides 3690-3706 and 4884-4900 in Fig. 2(underlines representing base substitutions) occur in the sequenced region, and could represent target site reduplications for other retroposition events in the history of this human DNA fragment. McBride *et al.* (1989) hypothesized that tRNA genes could be functional retrogenes, since short direct terminal repeat sequences flank many members of the dispersed tRNA gene family. The valine $tRNA_{AAC}$ gene and *Alu* member C are encompassed within a 1.8-kb DNA segment delimited by direct repeat III and also within a 1.2-kb DNA segment bounded by direct repeat II. The valine $tRNA_{CAC}$ and lysine tRNA genes are centered in a 1.5-kb DNA segment delimited by direct repeat I. Thus, it is conceivable that the valine $tRNA_{AAC}$ gene and *Alu* C in one case, and the valine $tRNA_{CAC}$ and lysine tRNA genes in the other case, were integrated into the human DNA segment in λhVKV7 by RNA-mediated retroposition events.

A variety of *cis*-acting structural motifs in the 5'- and even 3'-extragenic regions of genes transcribed by RNA polymerase III have been shown to regulate or modulate the transcriptional activities of those genes (reviewed in

Geiduschek and Tocchini-Valentini, 1988; Sollner-Webb, 1988). Several genes transcribed by RNA polymerase III have even been shown to have TATA-like promoter elements, including the genes encoding several tRNAs (Young *et al.*, 1986; Lee *et al.*, 1987; Capone, 1988), 5S rRNA (Tyler, 1987), 7SK RNA (Murphy *et al.*, 1987), and U6 RNA (Das *et al.*, 1988). Since the TATA sequences may be essential for transcription of some of these genes, it can be speculated that some type of interaction occurs between the components of the RNA polymerase II and III transcription systems (Söllner-Webb, 1988).

The two valine tRNA genes in λhVKV7 also have potential modulatory elements in their upstream flanking regions. Two presumptive TATA-like sequences occur from -32 to -26 and from -41 to -36 upstream from the valine $tRNA_{AAC}$ gene (5'-TATAAAA-3', residues 4651 to 4657, and 5'-TATAAC-3', residues 4642 to 4647, respectively, in Fig. 2). It can be speculated that the TATA-like promoter elements upstream from the valine $tRNA_{AAC}$ gene may be involved in the modulation of transcription for this gene, and that the transcription of the valine $tRNA_{AAC}$ gene may be modulated by some type of interaction between the *trans*-acting components of the RNA polymerase II and III transcription systems. The pentanucleotide sequence 5'-CAGGA-3' occurs three times (from -19 to -15, from -33 to -29, and from -77

to -73, residues 526-530, 512-516, and 368-372, respectively) upstream from the valine $tRNA_{CAC}$ gene, near an alternating purine-pyrimidine tract. An identical pentanucleotide sequence occurs from -46 to -42 upstream from a human valine $tRNA_{AAC}$ gene characterized by Arnold *et al.* (1986), and may be the core of an extragenic promoter element implicated in the high transcriptional efficiency of this valine tRNA gene (Arnold and Gross, 1987). From the results of the *in vitro* transcription assays (Fig. 4), it can be speculated that the three 5'-CAGGA-3' sequences upstream from the valine $tRNA_{CAC}$ gene in λhVKV7 have a similar modulatory function in the high transcriptional efficiency of this valine tRNA gene as it does for the valine $tRNA_{AAC}$ gene characterized by Gross and co-workers (Arnold *et al.*, 1986; Arnold and Gross, 1987; Arnold *et al.*, 1988). However, it is surprising that the same structural element could be implicated in the high transcriptional efficiencies of valine tRNA genes encoding both the major and minor valine tRNA isoacceptors. It is unclear at this time why the valine $tRNA_{AAC}$ gene in λhVKV7, encoding the major valine tRNA isoacceptor, has an apparently lower transcriptional efficiency than the valine $tRNA_{CAC}$ gene in λhVKV7, encoding the minor valine tRNA species.

CHAPTER 2

CHARACTERIZATION OF THE TRANSCRIPTION UNIT AND TWO PROCESSED PSEUDOGENES OF CHIMPANZEE TRIOSEPHOSPHATE ISOMERASE

Introduction

Triosephosphate isomerase (TPI, EC 5.3.11) catalyzes the interconversion of dihydroxyacetone phosphate and glyceraldehyde 3-phosphate in the glycolytic pathway. The catalytic efficiency of the enzyme is very high, increasing the rate of catalysis by approximately 10^{10} (Albery and Knowles, 1977). Its rate of catalysis approaches that of being diffusion-controlled, such that the enzyme has been termed the "perfect catalyst" (Albery and Knowles, 1977). The ubiquitous housekeeping enzyme, consisting of two identical polypeptide chains of 248 amino acids each in mammals (Corran and Waley, 1975; Lu *et al.*, 1984), has been highly conserved over evolutionary time. Studies on the genetic variability of the enzyme (Decker and Mohrenweiser, 1981; Asakawa and Iida, 1985) suggest that essentially any mutation affecting the protein is deleterious due to its high degree of perfection.

When hominoid (human, chimpanzee, gorilla, orangutan, and gibbon) cells are stimulated to proliferate, TPI

activity increases, and a second electrophoretically unique isozyme of TPI is expressed (Kester *et al.*, 1977; Rogers *et al.*, 1980; Rubinson *et al.*, 1971). In contrast, non-hominoid cells only synthesize the constitutive TPI subunit. Genetic studies have demonstrated that the proliferation-specific subunit of TPI (TPI-2) and the constitutive subunit (TPI-1) appear to be products of the same structural locus (Decker and Mohrenweiser, 1981; Asakawa and Iida, 1985). Both RNA and protein synthesis are apparently required for the expression of TPI-2 (Kester *et al.*, 1977). However, pulse-chase experiments did not demonstrate a precursor-product relationship between the two isozymes (Kester *et al.*, 1977). The information necessary for expression of TPI-2 has been proposed to be inherent in the TPI structural locus (Decker and Mohrenweiser, 1981; Asakawa and Iida, 1985).

Triosephosphate isomerase provides an interesting model for the study of gene evolution since it encodes an ancient, ubiquitous, well characterized, relatively small, highly conserved protein. The amino acid sequence for TPI has been determined directly or has been deduced from DNA sequences for twelve species: human (Lu *et al.*, 1984; Maquat *et al.*, 1985; Brown *et al.*, 1985), rhesus (Old and Mohrenweiser, 1988), rabbit (Corran and Waley, 1975), chicken (Furth *et al.*, 1974; Straus and Gilbert, 1985), coelacanth (Kolb *et al.*, 1974), maize (Marchionni and Gilbert, 1986),

Schizosaccharomyces pombe (Russell, 1985), *Saccharomyces cerevisiae* (Alber and Kawasaki, 1982), *Aspergillus nidulans* (McKnight *et al.*, 1986), *Bacillus stearothermophilus* (Artavanis-Tsakonas and Harris, 1980), *Escherichia coli* (Pichersky *et al.*, 1984), and *Trypanosoma brucei* (Swinkels *et al.*, 1986). The DNA sequences encoding the enzyme are known for nine of these species: human, rhesus, chicken, maize, *S. pombe*, *S. cerevisiae*, *A. nidulans*, *E. coli*, and *T. brucei*.

Studies on the sequences of the protein have shown a very high degree of conservation across prokaryotes and eukaryotes, indicating that all triosephosphate isomerase genes have arisen from a single primordial gene. A comparison of the primary structures of the protein from different species has resulted in an estimate of 2.8 mutations per 100 amino acid residues per 100 million years as the rate of acceptable amino acid substitutions, which is comparable with that of cytochrome c (Dayhoff,1978). Studies from different human populations have revealed that null alleles of TPI may occur in one of every 100-200 newborns (Asakawa and Mohrenweiser, 1982). These observations also suggest that virtually any mutation results in an enzyme which is less efficient.

The present study of the common chimpanzee (*Pan troglodytes*) TPI gene family was undertaken in order to obtain the complete nucleotide sequence, including the

introns, of a hominoid TPI gene capable of expressing both TPI isoenzymes, so as to possibly elucidate the mechanism responsible for production of the inducible TPI-2 subunit. In addition, the structural information obtained from this work should be useful in future phylogenetic comparisons. The characterization of the chimpanzee TPI transcription unit and two of its processed pseudogenes is presented. The functional gene was found to have a very high identity with the human TPI gene such that identical amino acid sequences are predicted for the two proteins. The upstream flanking region of one of the processed pseudogenes exhibits a high sequence identity with a human TPI processed pseudogene (Brown *et al.*, 1985), indicating that its insertion into the genome occurred prior to the divergence of the two species. Furthermore, a solitary LTR (long terminal repeat) that exhibits a high identity with human endogenous retroviral (HERV-K) LTRs was found to be integrated within the structural region of this processed retropseudogene.

Materials and Methods

Genomic Library Screening and Analysis of TPI Genomic Clones

A chimpanzee genomic library harbored in bacteriophage λ Charon 32 (Chang and Slightom, 1984) was screened by the plaque hybridization procedure (Benton and Davis, 1977). The recombinant λ phage were grown on stationary-phase *Escherichia coli* K-12 DP50*supF* in 2x NZCYM-DT medium (Blattner *et al.*, 1978; Maniatis *et al.*, 1982). The 1.2-kb *PstI* fragment corresponding to the human TPI cDNA insert was isolated from the recombinant plasmid pHTPI-5a (Maquat *et al.*, 1985), and the ^{32}P-labeled DNA fragments derived from nick-translation of this fragment (Rigby *et al.*, 1977) were utilized as hybridization probe. Nitrocellulose filters with bound DNA were prewashed, prehybridized, and hybridized as described in Maniatis *et al.* (1982). DNAs from the plaque-pure clones harboring presumptive members of the chimpanzee TPI gene family were analyzed by digestion with restriction endonucleases followed by electrophoresis on 1% agarose gels. The gels were stained with ethidium bromide and the banding patterns compared.

Nucleotide Sequence Analysis

DNA fragments encompassing the putative TPI gene and pseudogenes were subcloned into the polylinker region of

M13mp18 RF and/or M13mp19 RF, and transformed into competent *E. coli* JM101 (Yanisch-Perron *et al.*, 1985). A series of overlapping deletion subclones for DNA sequence analysis were generated for each of the subcloned fragments by the 3'-exonuclease activity of T4 DNA polymerase (Dale *et al.*, 1985). DNA sequence analysis was done by the dideoxynucleotide chain-termination procedure (Sanger *et al.*, 1980). Both strands of the DNA fragments encompassing the TPI structural regions were sequenced to ensure accuracy. The Klenow fragment of *E. coli* DNA polymerase I (New England Biolabs) or modified T7 DNA polymerase (Sequenase from U.S. Biochemical) with either [^{35}S]deoxyadenosine 5'-(α-thio)-triphosphate (New England Nuclear/Dupont) or [α-^{32}P]dATP (ICN), was utilized in the sequencing reactions for chain extension from a 17-bp oligodeoxynucleotide universal primer (New England Biolabs) complementary to a segment of the M13 vector sequence. In several cases, in order to overlap sequences between fragments, synthetic oligodeoxynucleotide primers obtained from New England Nuclear/Dupont were used instead of the universal sequencing primer. Ambiguities due to compression effects were resolved by utilizing the dGTP analog dITP (Barnes *et al.*, 1983). The sequencing gels were dried and autoradiographed using Kodak XAR-5 film. DNA sequences were stored and analyzed by Irma Pirtle with microcomputer

programs (Larson and Messing, 1983; Pearson and Lipman, 1988) and compared with other sequences from GenBank (Version 55.0). The percent divergences of the nucleotide sequences of the two processed pseudogenes from the functional gene were determined in order to estimate their approximate ages. Only coding region sequences were used in these determinations. The processed pseudogenes were assumed to mutate at a rate free from selective constraints (Graur *et al.*, 1989), and the functional gene coding region was assumed to be conserved, except for synonymous substitutions. Since synonymous changes in functional genes occur at approximately half the rate as base substitutions in pseudogenes (Li *et al.*, 1981; Miyata and Yasunaga, 1981), two-thirds of the synonymous changes were considered as changes in the processed pseudogenes following integration, and one-third were assumed to be due to evolutionary drift in the functional gene (Lee *et al.*, 1983). It was also assumed that, for neutral mutations, the rate of divergence is 0.7% per million years (Perler *et al.*, 1980; Ochman and Wilson, 1987).

Genomic DNA Preparation and Hybridization

Chimpanzee (*Pan troglodytes*) and human genomic DNAs were prepared from lymphocytes essentially by the method of Blin and Stafford (1976). Heparinized chimpanzee whole blood was obtained from the Yerkes Primate Research Center.

The genomic DNAs (10 μg) were digested with several restriction endonucleases (BRL) and the fragments were resolved on 0.8% agarose gels. Partially depurinated and denatured DNA was transferred (Southern, 1979) to positively-charged nylon membranes (Zeta-Probe from BioRad) in 20x SSC (0.3 M sodium citrate, 3.0 M NaCl, pH 7.0). The 5' portion of the human TPI cDNA insert of pHTPI-5a (Maquat *et al.*, 1985), obtained by digesting the purified 1.2-kb cDNA insert with *EcoRI* and isolating the 0.7-kb *PstI/EcoRI* fragment, encompassing most of the TPI coding region, was used to generate the probe for genomic blots. DNA fragments for probes were prepared by the random primers method (Feinberg and Vogelstein, 1983) with [α-^{32}P]dCTP (3000 Ci/mmole, ICN). Membranes were prehybridized overnight at 68°C in 1.5x SSPE (1x SSPE is 10 mM sodium phosphate, 180 mM NaCl, 1 mM EDTA, pH 7.4), 1.0% SDS, 0.5% lowfat dry milk, 10% dextran sulfate, and 0.5 mg/ml denatured salmon sperm DNA. After prehybridization, the ^{32}P-labeled fragments were added to the prehybridization solution, and hybridization was for 24 h. Following hybridization, the filters were washed at room temperature successively in (i) 2x SSC, 0.1% SDS for 30 min, (ii) 0.5x SSC, 0.1% SDS for 30 min, and (iii) 0.1x SSC, 0.1% SDS for 1 h, and autoradiographed.

Results and Discussion

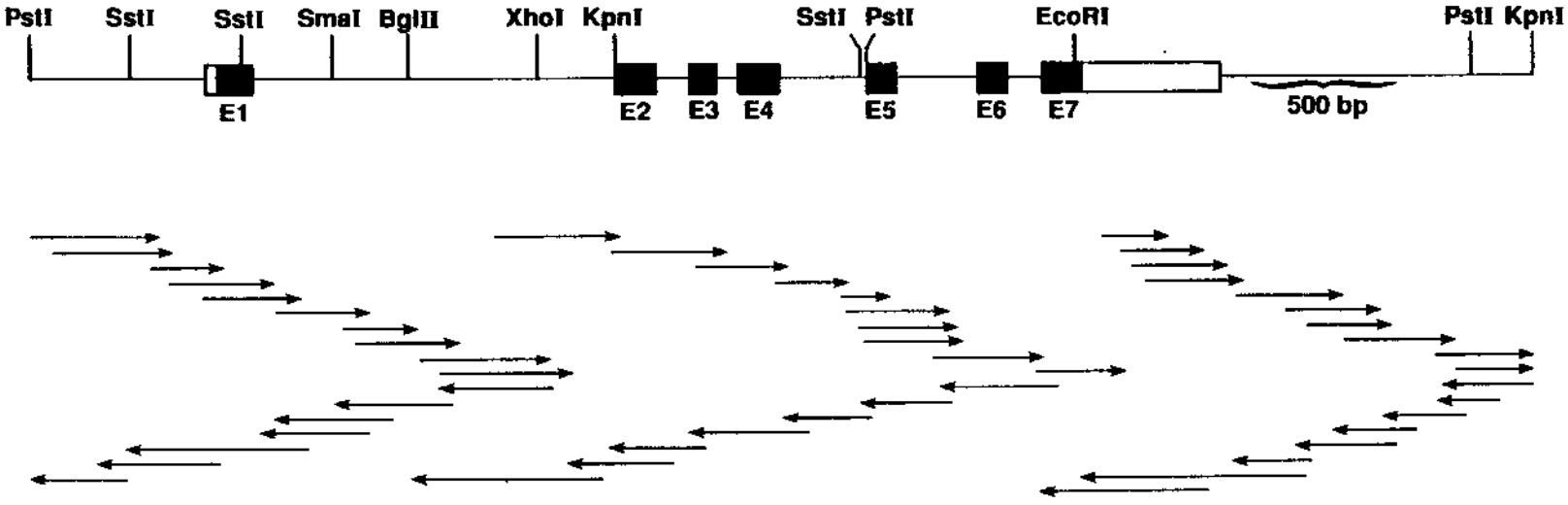
Isolation and Characterization of the TPI Clones

Twelve chimpanzee genomic clones were isolated which gave positive hybridization signals with the human TPI cDNA probe. Initially, the cloned DNA fragments were characterized by restriction endonuclease digestions and Southern blot hybridization analyses, and these 12 clones were found to correspond to three different sequences. The three distinct clones, subsequently designated λchpTPI-1, λψchpTPI-A, and λψchpTPI-B, were further analyzed by Southern blot hybridization analysis and DNA sequence analysis, which permitted the construction of the physical maps shown in Fig. 6. Specifically, two *PstI* fragments (2.3- and 2.8-kb) and an overlapping 3.0-kb *KpnI* fragment from λchpTPI-1, depicted in Fig. 6A, were found to hybridize with the human TPI cDNA probe. A 2.3-kb *XbaI/EcoRI* fragment and an overlapping 2.9-kb *SstI* fragment from λψchpTPI-A, shown in Fig. 6B, and a 3.1-kb *KpnI* fragment from λψchpTPI-B, depicted in Fig. 6C, were also found to hybridize with the human TPI cDNA probe. These hybridizing fragments and appropriate subfragments were subsequently sequenced by the dideoxynucleotide chain termination procedure, and the sequencing strategies are shown below the derived physical maps in Figs. 6A-6C. As discussed below, two of the clones

Fig. 6. Physical maps and DNA sequencing strategies of three segments of chimpanzee genomic DNA encompassing (A) the TPI transcription unit and (B-C) two TPI retropseudogenes. Black rectangles represent (A) the seven TPI exons and (B-C) the inactive coding regions. The open rectangles represent (A) the flanking untranslated regions or (B-C) the noncoding sequences between the direct repeats of the pseudogene in λψchpTPI-A and the 5'-flanking noncoding sequence in λψchpTPI-B. The stippled rectangle in (B) depicts the locus of an *Alu* element. The hatched rectangle in (C) represents the locus of a long terminal repeat (LTR). The arrays of arrows indicate the direction and extent of DNA sequences derived from the overlapping deletion subclones or primer-extended sequences obtained with synthetic oligodeoxynucleotides.

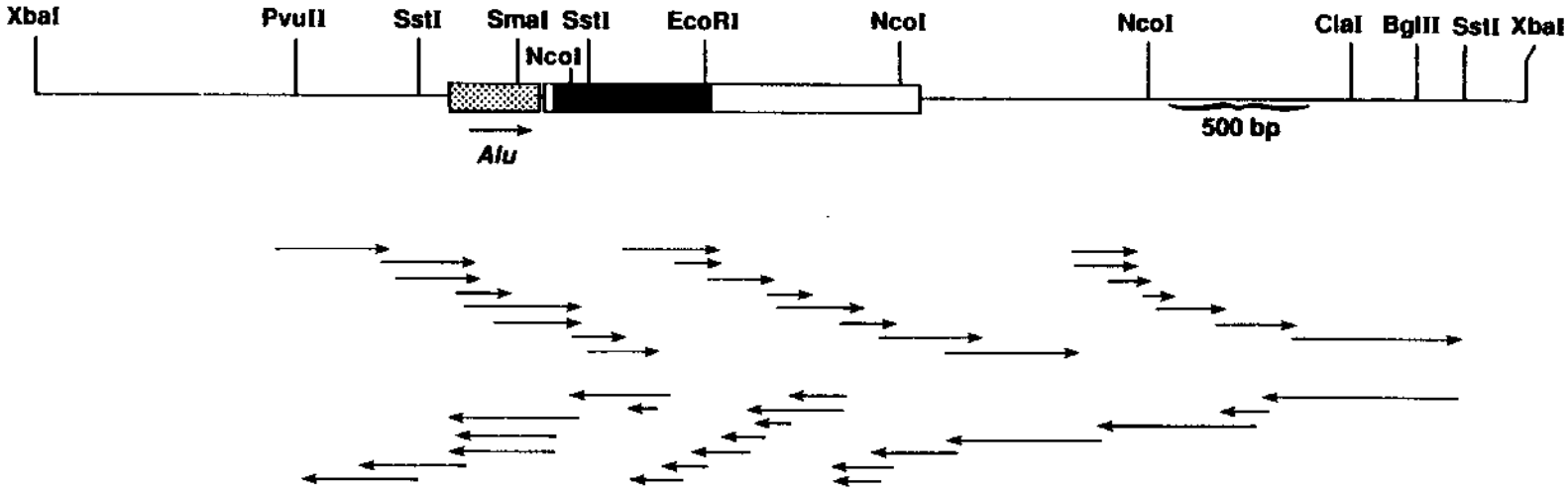
A

λchpTPI-1

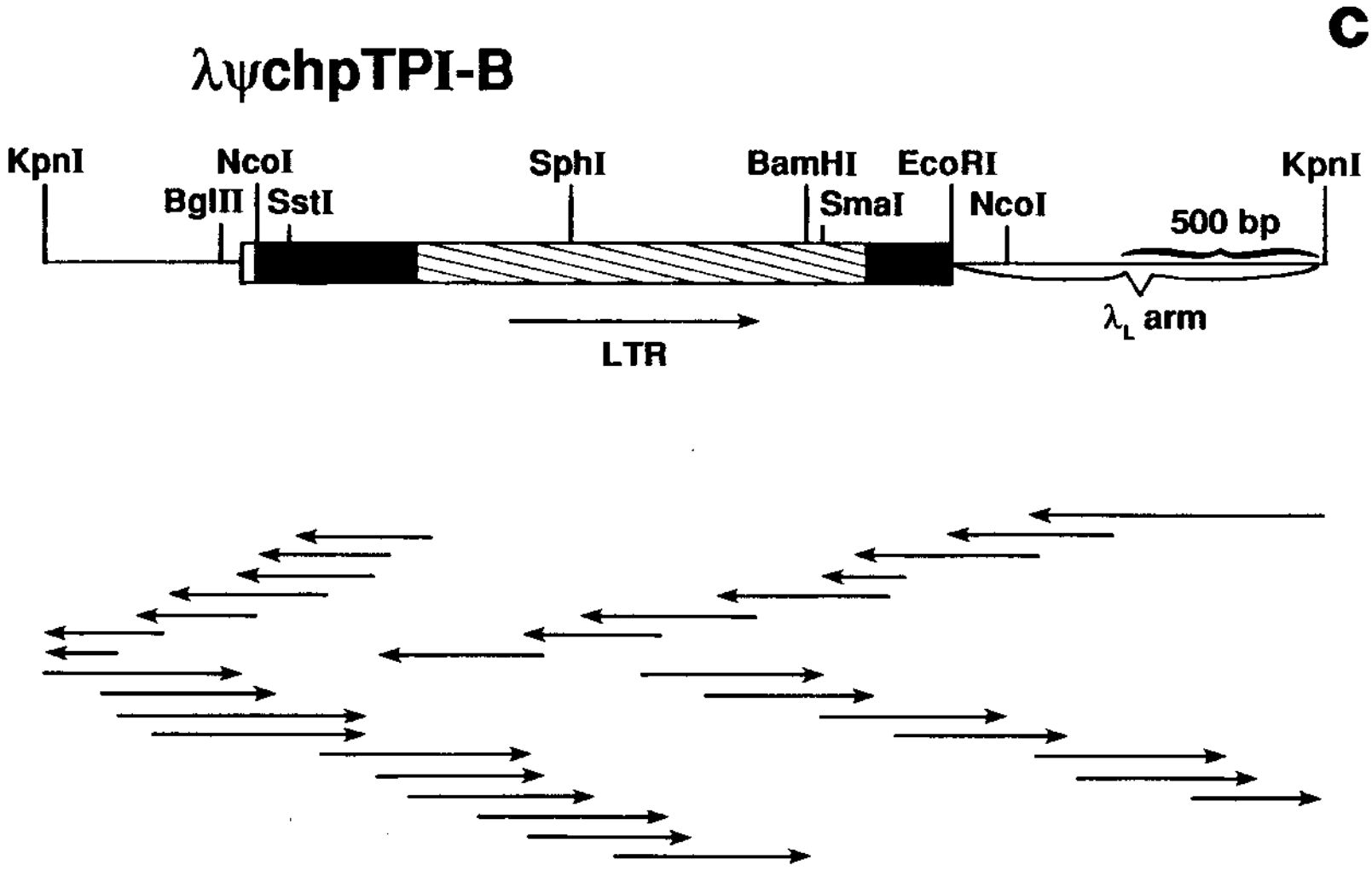

PstI
SstI
SstI
SmaI
BglII
XhoI
KpnI
SstI
PstI
EcoRI
PstI
KpnI
E1
E2
E3
E4
E5
E6
E7
500 bp

B

λψchpTPI-A


XbaI
PvuII
SstI
SmaI
SstI
NcoI
EcoRI
NcoI
NcoI
ClaI
BglII
SstII
XbaI
Alu
500 bp

c


λψchpTPI-B
KpnI
NcoI
BglII
SstI
SphI
BamHI
SmaI
EcoRI
NcoI
KpnI
500 bp
λL arm
LTR

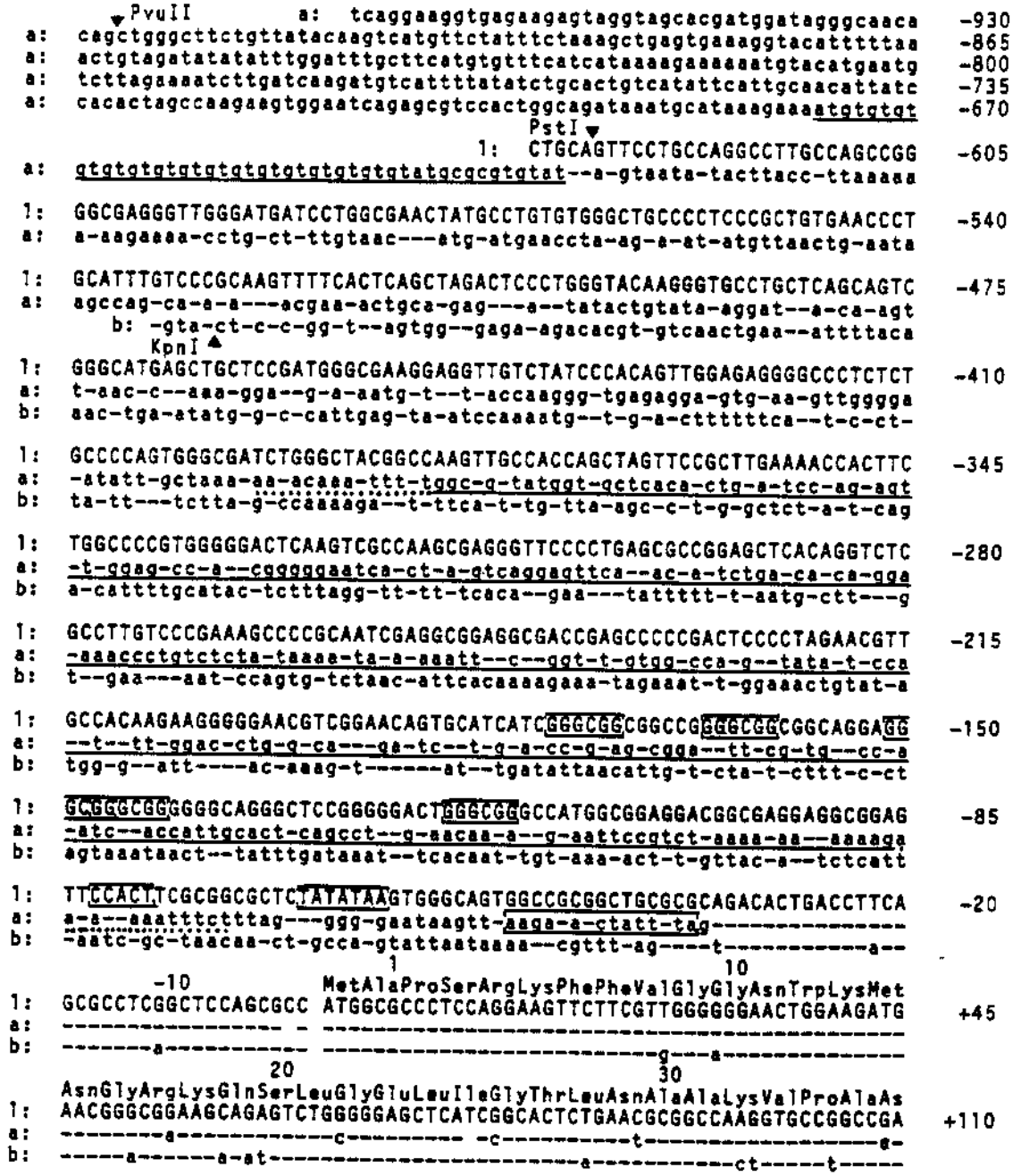
(λψchpTPI-A and λψchpTPI-B) were deduced to contain TPI processed pseudogenes since they had uninterrupted but defective open reading frames with great identity to human TPI gene sequences (Brown *et al.*, 1985). The existence of a functional TPI gene with six introns in λchpTPI-1 was indicated by a comparison of the sequence of hybridizing fragments from λchpTPI-1 with the two pseudogene sequences.

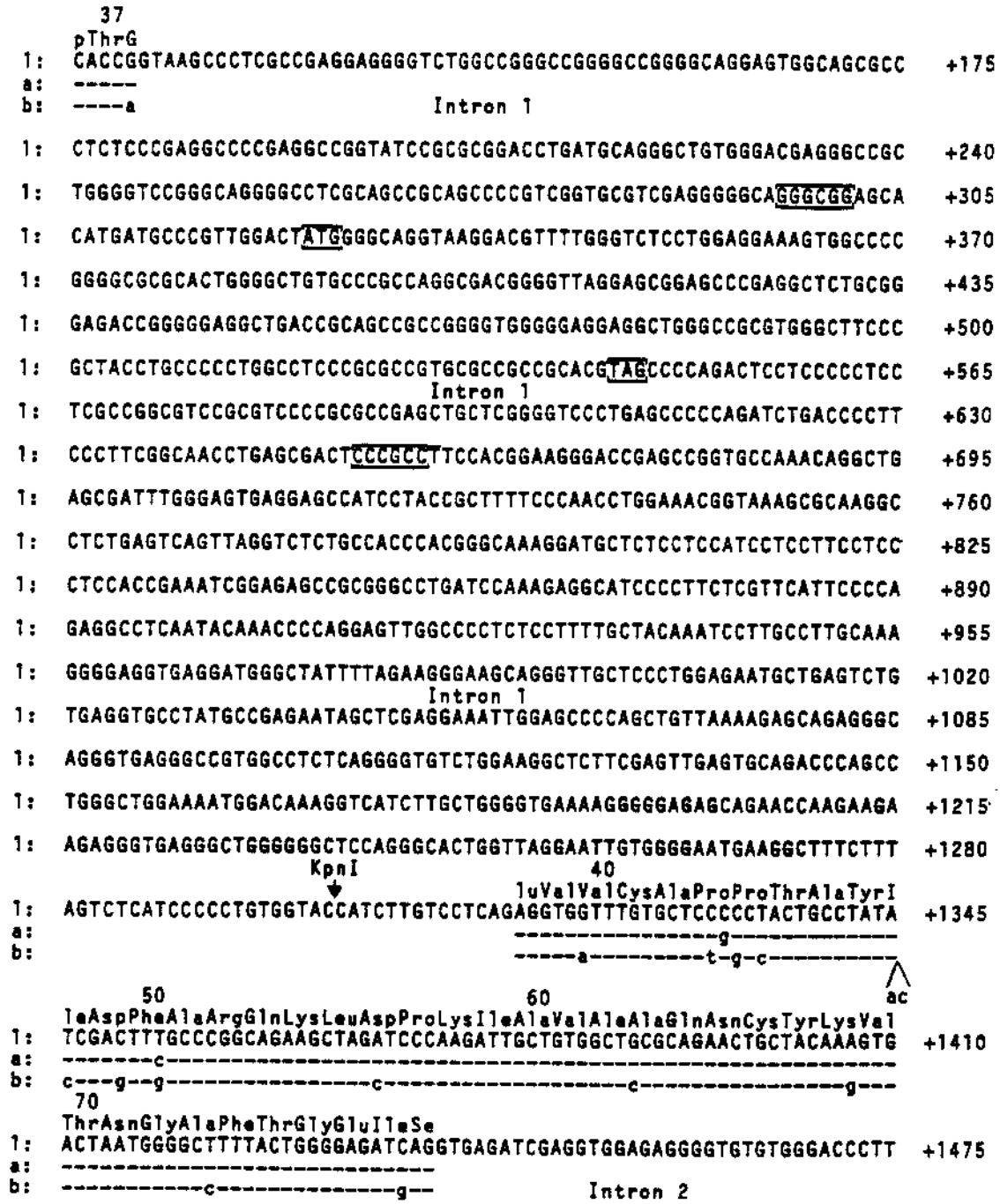
Nucleotide Sequence of the TPI Transcription Unit

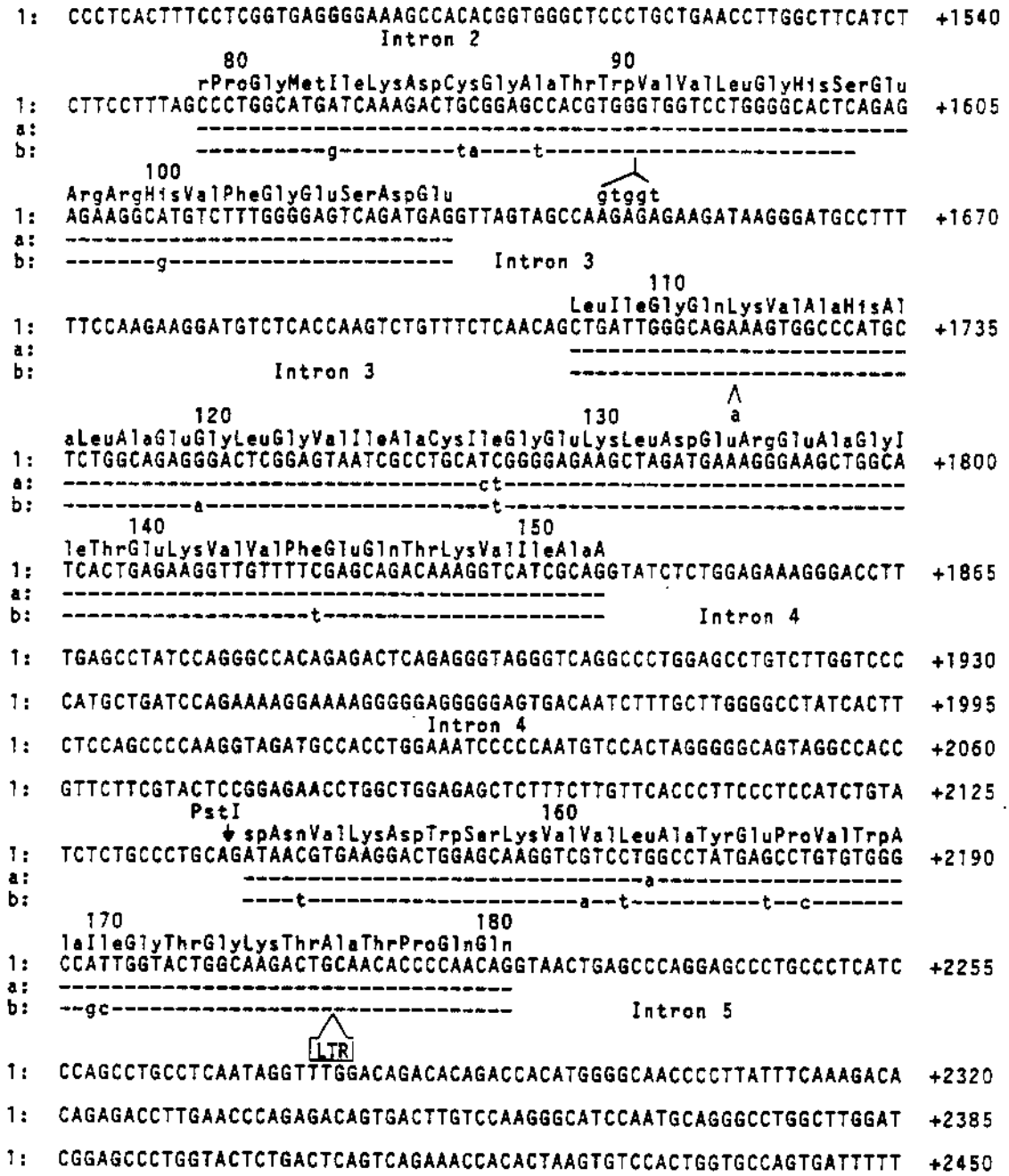
The 3.5-kb chimpanzee TPI transcription unit was found to be encompassed within the 4957-bp sequenced region of the genomic clone λchpTPI-1 shown in Fig. 7. The intron-exon boundaries in the functional gene shown in Fig. 7 were delineated by comparisons with the nucleotide sequences of the two chimpanzee processed pseudogenes (also shown in Fig. 7) described below, with the human TPI gene (Brown *et al.*, 1985), human TPI cDNA (Maquat *et al.*, 1985), and the rhesus TPI gene (Old and Mohrenweiser, 1988). Since the two chimpanzee processed pseudogenes represent cDNA retrosequences, although defective (Vanin, 1985; Hull and Will, 1989), the justification was made to regard them as *de facto* chimpanzee TPI cDNA clones in delineating the intron/exon boundaries of the cognate chimpanzee TPI gene. As shown in Fig. 7, the gene has seven exons and six introns, with no intron sequences in the untranslated

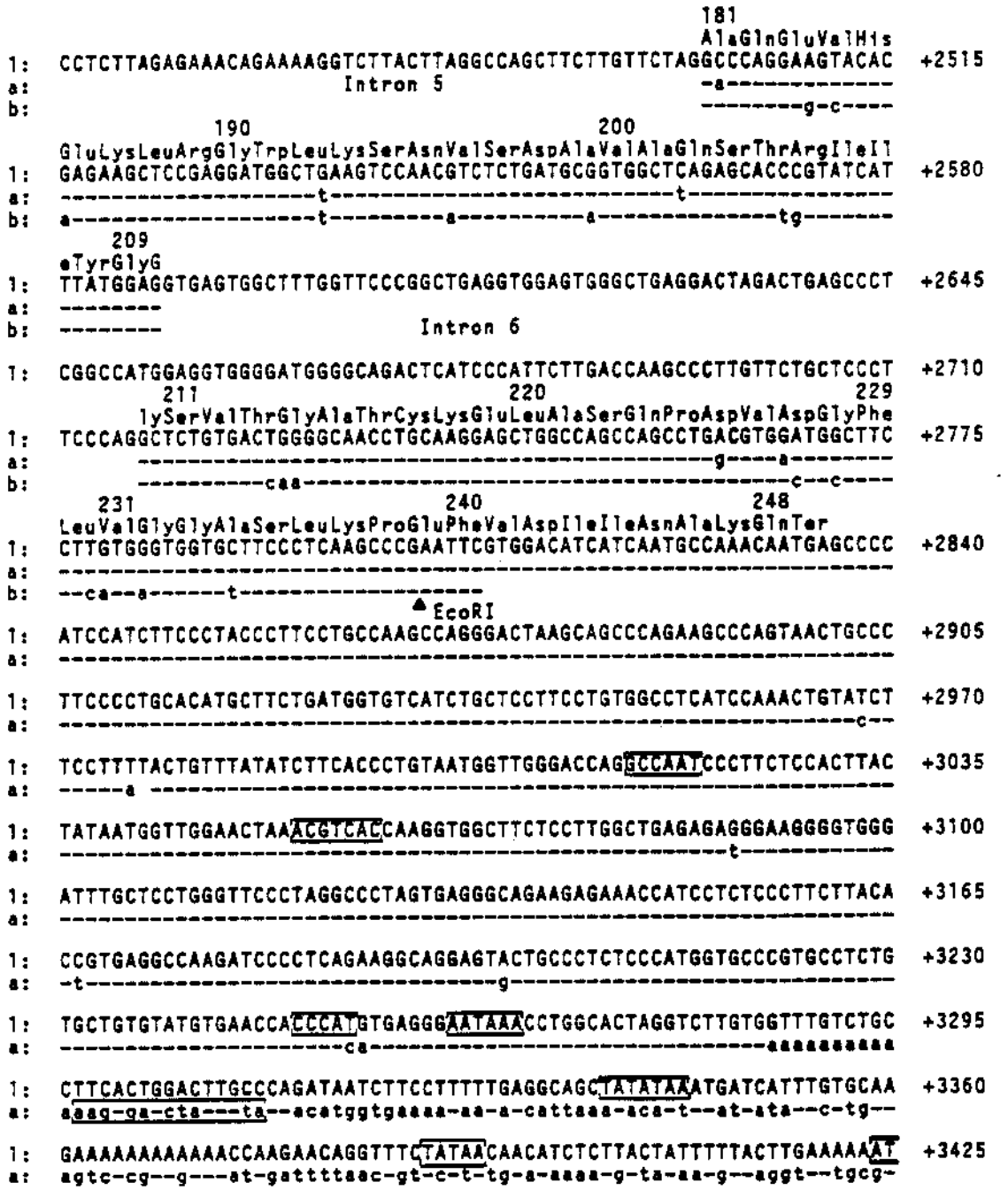
Fig. 7. Nucleotide sequences of the chimpanzee transcription unit and two TPI retropseudogenes. The sequence of a segment of the genomic clone λchpTPI-1 (denoted by 1 at the left-hand margin) is shown in upper-case letters. The sequences of the two processed pseudogenes derived from segments of the genomic clones λψchpTPI-A and λψchpTPI-B (denoted by a and b, respectively, in the left-hand margin) are shown in lower-case letters. The dashes in the sequences of the pseudogenes indicate sequence matches with the functional gene sequence. Mismatches and flanking sequences beyond those of the functional gene are shown in lower-case letters. Open arrows and lower-case letters represent insertions in the pseudogene sequences. Deletions in the processed pseudogenes are simply denoted by spacing. Closed arrows indicate restriction endonuclease cleavage sites. The deduced amino acid sequence of chimpanzee TPI is shown above the functional gene sequence. A long tract of nucleotides with the typical alternating purine-pyrimidine motif which has the potential to adopt the Z-DNA conformation in λψchpTPI-A is underlined. An *Alu* family member in the same clone is also underlined, with the relative polarity from 5' to 3'. The *Alu* direct repeat sequences are denoted by dotted lines. The direct repeat sequences of the pseudogene in λψchpTPI-A are boxed. The TATA, CCAAT, and GC

promoter elements and the polyadenylation signals for the functional gene, and the supposed promoter elements for the putative open reading frame in the TPI gene 3'-flanking region are boxed. The initiation and termination codons of the open reading frames are also boxed. The sequence of the indicated LTR insert in the pseudogene sequence from λψchpTPI-B is shown in Fig. 3. Note that the 5' direct repeat (AGACTG) which precedes the LTR is part of the coding region of the TPI pseudogene. The synthetic primers 5'-GGGACAAGGCGAGACCTG-3' (starting at -287), 5'-CCCTGCATCAGGTCCGCG-3' (at 205), 5'-CCGAGCAGCTCGGCGCGG-3' (at 584), 5'-GGCAGAGACCTAACTGAC-3' (at 767), and 5'-CCACAAGACCTAGTGCCAG-3' (at 3268), were used to complete the sequence of the functional gene. The primer 5'-CTCTGCTTTCGCCCGTTC-3' (starting at 45) was used to complete the sequence of the processed pseudogene from λψchpTPI-A.


PvuII
a: tcaggaaggtgagaagagtaggtagcacgatggatagggcaaca -930
a: cagctgggcttctgttatacaagtcatgttctatttctaaagctgagtgaaaggtacatttttaa -865
a: actgtagatatatatttggatttgcttcatgtgtttcatcataaaagaaaaaatgtacatgaatg -800
a: tcttagaaaatcttgatcaagatgtcatttttatatctgcactgtcatattcattgcaacattatc -735
a: cacactagccaagaagtggaatcagagcgtccactggcagataaatgcataaagaaaatgtgtgt -670
PstI
1: CTGCAGTTCCTGCCAGGCCTTGCCAGCCGG -605
a: gtgtgtgtgtgtgtgtgtgtgtgtgtatgcgcgtgtat--a-gtaata-tacttacc-ttaaaaa
1: GGCGAGGGTTGGGATGATCCTGGCGAACTATGCCTGTGTGGGCTGCCCCTCCCGCTGTGAACCCT -540
a: a-aagaaaa-cctg-ct-ttgtaac---atg-atgaaccta-ag-a-at-atgttaactg-aata
1: GCATTTGTCCCGCAAGTTTTCACTCAGCTAGACTCCCTGGGTACAAGGGTGCCTGCTCAGCAGTC -475
a: agccag-ca-a-a---acgaa-actgca-gag---a--tatactgtata-aggat--a-ca-agt
b: -gta-ct-c-c-gg-t--agtgg--gaga-agacacgt-gtcaactgaa--attttaca
KpnI
1: GGGCATGAGCTGCTCCGATGGGCGAAGGAGGTTGTCTATCCCACAGTTGGAGAGGGGCCCTCTCT -410
a: t-aac-c--aaa-gga--g-a-aatg-t--t-accaaggg-tgagagga-gtg-aa-gttgggga
b: aac-tga-atatg-g-c-cattgag-ta-atccaaaatg--t-g-a-ctttttca--t-c-ct-
1: GCCCCAGTGGGCGATCTGGGCTACGGCCAAGTTGCCACCAGCTAGTTCCGCTTGAAAACCACTTC -345
a: -atatt-gctaaa-aa-acaaa-ttt-tggc-g-tatggt-gctcaca-ctg-a-tcc-ag-agt
b: ta-tt---tctta-g-ccaaaaga--t-ttca-t-tg-tta-agc-c-t-g-gctct-a-t-cag
1: TGGCCCCGTGGGGGACTCAAGTCGCCAAGCGAGGGTTCCCCTGAGCGCCGGAGCTCACAGGTCTC -280
a: -t-ggag-cc-a--cgggggaatca-ct-a-gtcaggagttca--ac-a-tctga-ca-ca-gga
b: a-cattttgcatac-tctttagg-tt-tt-tcaca--gaa---tattttt-t-aatg-ctt---g
1: GCCTTGTCCCGAAAGCCCCGCAATCGAGGCGGAGGCGACCGAGCCCCCGACTCCCCTAGAACGTT -215
a: -aaaccctgtctcta-taaaa-ta-a-aaatt--c--ggt-t-gtgg-cca-g--tata-t-cca
b: t--gaa---aat-ccagtg-tctaac-attcacaaaagaaa-tagaaat-t-ggaaactgtat-a
1: GCCACAAGAAGGGGGAACGTCGGAACAGTGCATCATCGGGCGGCGGCCGGGGCGGCGGCAGGAGG -150
a: --t--tt-ggac-ctg-g-ca---ga-tc--t-g-a-cc-g-ag-cgga--tt-cg-tg--cc-a
b: tgg-g--att----ac-aaag-t------at--tgatattaacattg-t-cta-t-cttt-c-ct
1: GCGGGCGGGGGGGCAGGGCTCCGGGGGACTGGGCGGGCCATGGCGGAGGACGGCGAGGAGGCGGAG -85
a: -atc--accattgcact-cagcct--g-aacaa-a--g-aattccgtct-aaaa-aa--aaaaga
b: agtaaataact--tatttgataaat--tcacaat-tgt-aaa-act-t-gttac-a--tctcatt
1: TTCCACTTCGCGGCGCTCTATATAAGTGGGCAGTGGCCGCGGCTGCGCGCAGACACTGACCTTCA -20
a: a-a--aaatttctttag---ggg-gaataagtt-aaga-a-ctatt-tag----------
b: -aatc-gc-taacaa-ct-gcca-gtattaataaaa--cgttt-ag----t----------a--
1 10
-10 MetAlaProSerArgLysPhePheValGlyGlyAsnTrpLysMet
1: GCGCCTCGGCTCCAGCGCC ATGGCGCCCTCCAGGAAGTTCTTCGTTGGGGGGAACTGGAAGATG +45
a: ---------- - ----------
b: -------a---------- ----------g---a----------
20 30
AsnGlyArgLysGlnSerLeuGlyGluLeuIleGlyThrLeuAsnAlaAlaLysValProAlaAs
1: AACGGGCGGAAGCAGAGTCTGGGGGAGCTCATCGGCACTCTGAACGCGGCCAAGGTGCCGGCCGA +110
a: --------a----------c--------- -c----------t----------a-
b: -----a------a-at----------a----------ct-----t-----


37
pThrG
1: CACCGGTAAGCCCTCGCCGAGGAGGGGTCTGGCCGGGCCGGGGCCGGGGCAGGAGTGGCAGCGCC +175
a: -----
b: ----a Intron 1
1: CTCTCCCGAGGCCCCGAGGCCGGTATCCGCGCGGACCTGATGCAGGGCTGTGGGACGAGGGCCGC +240
1: TGGGGTCCGGGCAGGGGCCTCGCAGCCGCAGCCCCGTCGGTGCGTCGAGGGGGCAGGGCGGAGCA +305
1: CATGATGCCCGTTGGACTATGGGGCAGGTAAGGACGTTTTGGGTCTCCTGGAGGAAAGTGGCCCC +370
1: GGGGCGCGCACTGGGGCTGTGCCCGCCAGGCGACGGGGTTAGGAGCGGAGCCCGAGGCTCTGCGG +435
1: GAGACCGGGGGAGGCTGACCGCAGCCGCCGGGGTGGGGGAGGAGGCTGGGCCGCGTGGGCTTCCC +500
1: GCTACCTGCCCCCTGGCCTCCCGCGCCGTGCGCCGCCGCACGTAGCCCCAGACTCCTCCCCCTCC +565
Intron 1
1: TCGCCGGCGTCCGCGTCCCCGCGCCGAGCTGCTCGGGGTCCCTGAGCCCCCAGATCTGACCCCTT +630
1: CCCTTCGGCAACCTGAGCGACTCCCGCCTTCCACGGAAGGGACCGAGCCGGTGCCAAACAGGCTG +695
1: AGCGATTTGGGAGTGAGGAGCCATCCTACCGCTTTTCCCAACCTGGAAACGGTAAAGCGCAAGGC +760
1: CTCTGAGTCAGTTAGGTCTCTGCCACCCACGGGCAAAGGATGCTCTCCTCCATCCTCCTTCCTCC +825
1: CTCCACCGAAATCGGAGAGCCGCGGGCCTGATCCAAAGAGGCATCCCCTTCTCGTTCATTCCCCA +890
1: GAGGCCTCAATACAAACCCCAGGAGTTGGCCCCTCTCCTTTTGCTACAAATCCTTGCCTTGCAAA +955
1: GGGGAGGTGAGGATGGGCTATTTTAGAAGGGAAGCAGGGTTGCTCCCTGGAGAATGCTGAGTCTG +1020
Intron 1
1: TGAGGTGCCTATGCCGAGAATAGCTCGAGGAAATTGGAGCCCCAGCTGTTAAAAGAGCAGAGGGC +1085
1: AGGGTGAGGGCCGTGGCCTCTCAGGGGTGTCTGGAAGGCTCTTCGAGTTGAGTGCAGACCCAGCC +1150
1: TGGGCTGGAAAATGGACAAAGGTCATCTTGCTGGGGTGAAAAGGGGGAGAGCAGAACCAAGAAGA +1215
1: AGAGGGTGAGGGCTGGGGGGCTCCAGGGCACTGGTTAGGAATTGTGGGGAATGAAGGCTTTCTTT +1280
KpnI 40
luValValCysAlaProProThrAlaTyrI
1: AGTCTCATCCCCCTGTGGTACCATCTTGTCCTCAGAGGTGGTTTGTGCTCCCCCTACTGCCTATA +1345
a: ----------g----------
b: -----a---------t-g-c----------
ac
50 60
leAspPheAlaArgGlnLysLeuAspProLysIleAlaValAleAlaGlnAsnCysTyrLysVal
1: TCGACTTTGCCCGGCAGAAGCTAGATCCCAAGATTGCTGTGGCTGCGCAGAACTGCTACAAAGTG +1410
a: -------c----------
b: c---g--g----------c----------c----------g---
70
ThrAsnGlyAlaPheThrGlyGluIleSe
1: ACTAATGGGGCTTTTACTGGGGAGATCAGGTGAGATCGAGGTGGAGAGGGGTGTGTGGGACCCTT +1475
a: ----------
b: ----------c----------g-- Intron 2


Intron 2
Intron 3
Intron 4
Intron 5
LTR
PstI


181
AlaGlnGluValHis
1: CCTCTTAGAGAAACAGAAAAGGTCTTACTTAGGCCAGCTTCTTGTTCTAGGCCCAGGAAGTACAC +2515
a: Intron 5 -a----------
b: --------g-c----
190 200
GluLysLeuArgGlyTrpLeuLysSerAsnValSerAspAlaValAlaGlnSerThrArgIleIl
1: GAGAAGCTCCGAGGATGGCTGAAGTCCAACGTCTCTGATGCGGTGGCTCAGAGCACCCGTATCAT +2580
a: ----------t----------t----------
b: a----------t---------a----------a----------tg-------
209
eTyrGlyG
1: TTATGGAGGTGAGTGGCTTTGGTTCCCGGCTGAGGTGGAGTGGGCTGAGGACTAGACTGAGCCCT +2645
a: --------
b: -------- Intron 6
1: CGGCCATGGAGGTGGGGATGGGGCAGACTCATCCCATTCTTGACCAAGCCCTTGTTCTGCTCCCT +2710
211 220 229
lySerValThrGlyAlaThrCysLysGluLeuAlaSerGlnProAspValAspGlyPhe
1: TCCCAGGCTCTGTGACTGGGGCAACCTGCAAGGAGCTGGCCAGCCAGCCTGACGTGGATGGCTTC +2775
a: ----------g----a--------
b: ----------caa----------c--c----
231 240 248
LeuValGlyGlyAlaSerLeuLysProGluPheValAspIleIleAsnAlaLysGlnTer
1: CTTGTGGGTGGTGCTTCCCTCAAGCCCGAATTCGTGGACATCATCAATGCCAAACAATGAGCCCC +2840
a: ----------
b: --ca--a------t----------
▲EcoRI
1: ATCCATCTTCCCTACCCTTCCTGCCAAGCCAGGGACTAAGCAGCCCAGAAGCCCAGTAACTGCCC +2905
a: ----------
1: TTCCCCTGCACATGCTTCTGATGGTGTCATCTGCTCCTTCCTGTGGCCTCATCCAAACTGTATCT +2970
a: ----------c--
1: TCCTTTTACTGTTTATATCTTCACCCTGTAATGGTTGGGACCAGGCCAATCCCTTCTCCACTTAC +3035
a: -----a ----------
1: TATAATGGTTGGAACTAAACGTCACCAAGGTGGCTTCTCCTTGGCTGAGAGAGGGAAGGGGTGGG +3100
a: ----------t----------
1: ATTTGCTCCTGGGTTCCCTAGGCCCTAGTGAGGGCAGAAGAGAAACCATCCTCTCCCTTCTTACA +3165
a: ----------
1: CCGTGAGGCCAAGATCCCCTCAGAAGGCAGGAGTACTGCCCTCTCCCATGGTGCCCGTGCCTCTG +3230
a: -t----------g----------
1: TGCTGTGTATGTGAACCACCCATGTGAGGGAATAAACCTGGCACTAGGTCTTGTGGTTTGTCTGC +3295
a: ----------ca----------aaaaaaaaaa
1: CTTCACTGGACTTGCCCAGATAATCTTCCTTTTTGAGGCAGCTATATAAATGATCATTTGTGCAA +3360
a: aaag-ga-cta---ta--acatggtgaaaa-aa-a-cattaaa-aca-t--at-ata--c-tg--
1: GAAAAAAAAAAAACCAAGAACAGGTTTCTATAACAACATCTCTTACTATTTTTACTTGAAAAAAT +3425
a: agtc-cg--g---at-gattttaac-gt-c-t-tg-a-aaaa-g-ta-aa-g--aggt--tgcg-

```
1:  GTTTTGCGTAGCAGACTGTCATAGCCTTGAATGCCGGCTCCCTTTCTTTCTCCCTCCGAGTGGCT  +3490
a:  a-g-gaat----t-ga---a-ccatt-cacgatgtatacata----aaaacat-atgttt-ata-

1:  CTGGGGCTGTTGATTTCCCCAGAGCTTGGGTTGGGGTAGGGGCTCAGCCTCACCAGCTTTCAGCA  +3555
a:  gataactgta-ata-atgttttt-tcaat-aa-aaaataaatt-a-aagaggaa-ataaca-aac

1:  GCTGGTCTAGGCCAGCAGTGCCTCCCCACCTCCCCAAGGGGAGGGGTGGTGGCAAGGCCTCTGTG  +3620
a:  aagaaaaa-a-gatat-accaga-aaa-tg-taaa-gtttat-t-t-a-a--tttactagggcct

1:  CACAGTCTGTGGTATCACAGGGCTCACTGGTAGAGCAAGTAGCGCTTCATGGCAGGGGGCAAGGG  +3685
a:  tg---ataa-t---aa-a-tt-tag-aacaacataa-ca-tcat-a-t-ga-attt-att--aa-

1:  CAGGGCAGATACCTGGCCGAGCCGGGTATCCCCCAGGGTGTGGCGCACACACAGGCGGCTCAGGT  +3750
a:  ataa---a--c-a-a-aaacagaaa-c-gag-at--atgagtag-ag-tgggg-aa--gag-aa-

1:  GCAGAAGGGAGTGTGGCTCCGCTGGGAGAGAGAAGGAGGGGAATGTAAGTATGGGTGCAGCCACC  +3815
a:  -gga--t-actgc-aatcagtgca-c-tttctttt-g-aatg---a--a---tttaaa-ttagat

1:  AGCCAGATGTCCTCAAACTACAGGGTCCTACTCAGATGCCTTTCTGCTTTCCTGCTTTGAGTGTG  +3880
a:  t-ttgtgatagt-gc-caatatattaaaac-att--gttg-acaaaa-ga-aat-aaatt-g---

1:  CCCACCTTGGCTGAAAGGGGAATTTGAGATACCCGGAAGTTCTGCCTCCCAGATAAGATTTCACA  +3945
a:  tgg-agagacaactttaaaag--agag-gagttg-a--t--gctt--tta---a---taaaacgt

1:  CAGCCCTAGTCAGAGCTGGGGGTGAAGAGCTGGCTAAGGCCCTCTAAGCGACAGGCCAAGGTGGC  +4010
a:  a-aattgtaaa-agcacataca--t--tttat--aggaatgaaaggttgt-atttaacccagcag

1:  TCTGACAGTGGTGGAGCTGGCCCAGGCTTTGACTCCAGAGGCTTGGGAGCTGGGGCTGAGGTGAG  +4075
a:  t-ccca-cctt-ttgt-attagg--ctgg-tt-atgga--acaattttt-cat--a--g--gc--
                                      PstI▼
1:  GAGGGATGGGCCCTCCACTCTACAGCCCACACAACTGCAGAGAGCAGCTCCAAGCCCTGGACCCA  +4140
a:  -g-atgg--agtt--aggatg-a-ct-tt-cacctca--tcatcag--att-gatt--cataaa-

1:  GTCAGTTCCTGGGGAGGCTCCTCCCCTGCTGCCCCACCCTAAGGCCCTGCCTCCTCCACTGCTCT  +4205
a:  -a--tgaaaccta--tc-ct-a-atgca-a-tttacaatagt-tt-tct-t-ttatg-gaatcta

1:  CCTCCCTGGTGCCCAGGGCCCCAGTGTCTCCATCCTGAGGTGTGGCTGAGGAAGGAAGTAGGTAT  +4270
a:  atg---a--gtgatct-a--a----t----agg-actctcacccagc-ctccctt-ctgctc-gc
                                                  KpnI▼
1:  GTAGCACAGAGACAGGTTAGAGCCCAGGGAATCCGGTATACAGCCTGGGTACC              +4335
a:  taccaggttccta-caggcc-cagact-t-tagtc-gcgg-ccagc--aggaggggtgggggggg

a:  tgggggctgctttaatcagtacatccatatacaaatgaaagtcttgaatttatattaaaaatcta  +4400
a:  ggcaaaacattgtgcatgtgtatgttttaagttaaaatccctaatatacttaaaagtctgtttct  +4465
a:  atgatttcctgcctcactaaacttttcaagtaaccaaccaaatattaataataatgagcacatt  +4530
a:  agatagaaggcttctatggtatttcttcactcagttaaactaaaatttagtttcttttcttttc  +4595
a:  cgttacccagtgaatggcagcttgtgaaaaaccaaacatgaatagcttatagagataatgcctgt  +4660
a:  gtatagtgagagtaaaaattttaagtctgttaacaaatatgatttttgtgaatgaatacctaact  +4725
a:  tttgaatatcatgtctatttaatcgatgtttactattgtgacagagactgttttacttttctcag  +4790
a:  aatgaaggctagaaaagattttccaattgccctgtaactggatatggccatatgactaaattttg  +4855
a:  gccactgaaatatgagtggaaatggtgccacaacgtttaggcttgttaaaaaaaaaaatacaaaaa  +4920
a:  ccaagaaacacacaaaaaaactaccctaaaatcctcccaggtaagatcttccatgatacactctc  +4985
a:  catcctgctgactcaaatggaaatgaccgtaagtggaacctgagagccaaggtgaaggcagtggagctc
                                                                SstI▲
```

regions.

The chimpanzee TPI gene shares a great degree of identity with other sequenced vertebrate TPI genes (Brown *et al.*, 1985; Old and Mohrenweiser, 1988; Straus and Gilbert, 1985), with the juxtapositions of the exon/intron junctions being identical. In the chicken and other vertebrate TPI transcription units, all the introns fall either between or near the ends of secondary units of the repeating $\alpha\beta$ barrel motif of the protein (Straus and Gilbert, 1985). The locations of five of the eight maize TPI introns are identical to those in the vertebrate TPI genes, whereas the maize intron corresponding to the vertebrate intron 5 was frameshifted by three codons (Marchionni and Gilbert, 1986). The relative position of only intron 3 of the TPI gene between the filamentous fungi *A. nidulans* and vertebrates has been preserved throughout evolution (McKnight *et al.*, 1986). These observations allowed Marchionni and Gilbert (1986) to speculate that intron positions in ancient genes such as the TPI gene were conserved between distant species such as maize and vertebrates, and that the introns were in place prior to the plant-animal divergence, such that the ancestral TPI gene had at least eight introns, two of which were lost in the animal lineage. A comparison of the nucleotide sequences of the rhesus TPI gene introns 2-6 (Old, 1988) with the chimpanzee TPI gene introns 2-6

indicates that they differ from 1.4% to 9.9%, with an overall divergence of 9.5%, calculated according to Miyamoto *et al.* (1987). The chimpanzee and rhesus intron 1 sequences have an 87% identity, discounting a 158-bp deletion in the rhesus sequence (Old, 1988) corresponding to residues 272-429 in the chimpanzee sequence (Fig. 7).

It should be mentioned that a presumptive uninterrupted open reading frame (ORF) of 74 codons (from +324 to +545 in Fig. 7) occurs in intron 1 of the chimpanzee TPI gene, one nucleotide out of frame with the TPI gene. A corresponding ORF does not occur in the rhesus intron 1, since a portion of this region is deleted in the rhesus intron 1 (Old, 1988). It is rather doubtful that this ORF could be expressed, since no apparent promoter elements and polyadenylation signals occur in the nearby flanking regions. However, two consensus GC boxes (Kadonaga *et al.*, 1986) occur in intron 1, as depicted in Fig. 7.

The amino acid sequence of the chimpanzee TPI subunit deduced from the gene coding region comprises 248 amino acids, similar to the chain lengths of the other mammalian TPI polypeptides. Indeed, there are no differences between the predicted amino acid sequences of the chimpanzee and human TPI polypeptides. (Differences in thermostability between erythrocyte TPI from chimpanzee and human have been previously noted (Naidu *et al.*, 1984). In these studies, it was shown that the half lives for human and chimpanzee TPI

at 58°C are 37 min and 21 min, respectively. Since the genes coding for TPI in these two species predict an identical amino acid sequence, it is interesting to speculate that a species-specific modification of TPI might occur which could give rise to differences in thermostability.) There are only five nucleotide differences within the 750-bp coding regions of the chimpanzee and human TPI genes, all occurring in the codon wobble position (at nucleotides 33, 1353, 1596, 1769, and 2763, shown in Fig. 7). It should be noted that the Gly codon 94 is GGG in the chimpanzee and human genes, but GGA in the human cDNA, and the Leu codon 162 is CTG in the chimpanzee gene and human cDNA, but CTC in the human gene. There are 20 nucleotide differences between the chimpanzee and the rhesus coding regions (Old and Mohrenweiser, 1988), resulting in two amino acid differences (Ser 20 in chimpanzee to Asn 20 in rhesus and Asp 198 in chimpanzee to Glu 198 in rhesus). Ser is the consensus amino acid at position 20 of TPI polypeptides among species thus far examined, with the only exceptions being Asn in rabbit (Corran and Waley, 1975) and rhesus (Old and Mohrenweiser, 1988), Met in *E. coli* (Pichersky *et al.*, 1984), Gln in maize (Marchionni and Gilbert, 1986), and Glu in *B. stearothermophilus* (Artavanis-Tsakonas and Harris, 1980). Variability at amino acid 198 is greater, *e.g.* Glu (coelacanth and rhesus), Ala (*Trypanosome brucei*, *A.*

nidulans, *S.pombe*), Pro (maize), and Asp (*S. cerevisiae*, chicken, chimpanzee, human, and rabbit), with no general consensus amino acid residue.

The exon sequences of the chimpanzee TPI gene have a 79% identity to those of the chicken TPI gene (Straus and Gilbert, 1985), and there is a 90% identity (222 of the 248 amino acids in the chimpanzee sequence) between the amino acid sequences encoded by these genes. Ser 3 in the deduced chimpanzee amino acid sequence is missing in the deduced chicken amino acid sequence (Straus and Gilbert, 1985), as well as in the amino acid sequences of *A. nidulans* (McKnight *et al.*, 1986), *S. pombe* (Russell, 1985), *S. cerevisiae* (Alber and Kawasaki, 1982), and coelacanth (Kolb *et al.*, 1974). However, Ser 3 is present in rabbit (Corran and Waley, 1975), rhesus (Old and Mohrenweiser, 1988), and human (Lu *et al.*, 1984; Maquat *et al.*, 1985) amino acid sequences.

The Flanking Regions of the Transcription Unit

In the 5'-flanking region of the TPI gene, the sequence of 634-bp upstream from the ATG initiation codon was determined. Since the 5'-flanking direct repeat of retrosequences frequently delineates the 5'-residue of the mRNA, it can be inferred that the transcription start site of the chimpanzee TPI gene occurs 38 nucleotides upstream from the initiation codon, by a comparison with the upstream sequences flanking the two chimpanzee TPI retropseudogenes

shown in Fig. 7. A number of promoter element motifs identified as being both housekeeping and facultative gene promoters for RNA polymerase II-directed transcription (Serfling *et al.*, 1985; Martini *et al.*, 1986; Boyer *et al.*, 1989; Mitchell and Tjian, 1989) occur within the 596-bp sequence upstream from the assumed transcription start site in the chimpanzee TPI gene. The promoter elements potentially designated as being facultative gene promoter regions include a TATA sequence (TATATAA at -66 in Fig. 7, or 28 residues prior to the tentative transcription start site) and a CCAAT sequence (CCACT at -82, or 44 nucleotides from the transcription start site). Akin to housekeeping gene promoter regions, which generally have a high GC content with the GC box motif for binding the transcription factor Sp1 (Kadonaga *et al.*, 1986), the upstream region of the TPI gene is 91% GC, with five canonical GGGCGG hexanucleotide motifs from -177 to -115 in Fig. 7. Thus, the presumptive chimpanzee TPI gene promoter motif has characteristics of both facultative and housekeeping gene promoter regions, with the TATA and GC motifs, as well as a near consensus CCAAT element. The human carbonic anhydrase II gene also has a variety of promoter element motifs which exhibit characteristics of facultative and housekeeping gene regulatory elements (Shapiro *et al.*, 1987).

In an elegant study, Boyer *et al.* (1989) found that the

major determinants of the human TPI promoter region occur within the 131 nucleotides prior to the transcription start site. The chimpanzee and human 5'-flanking regions have a 98% identity between -634 and -39, with both having five GC elements and the TATA sequence. Thus, a strong argument can be made that the upstream regions flanking both primate genes should regulate TPI gene expression to essentially the same extent. Since a characteristic of housekeeping gene promoter regions is the presence of GC boxes and the lack of a TATA box (Dynan, 1986), the chimpanzee and human TPI gene upstream regions are exceptional in that they contain a TATA element normally associated only with facultative genes.

In the TPI gene 3'-flanking region shown in Fig. 7, the sequence of 1488 bp downstream from the TGA termination codon was determined. A presumptive AATAAA polyadenylation cleavage site occurs 426 nucleotides downstream from the termination codon, with the polyadenylation site located 19 residues downstream from the polyadenylation signal, deduced from the presence of an oligo(A) stretch in the sequence of the processed pseudogene in λψchpTPI-A (at nucleotide +3285 in Fig. 7). The 3'-terminus of the chimpanzee TPI gene has 98% identity with the sequence of 721 bp at the 3'-end of the human TPI gene (Brown *et al.*, 1985) and a 93% identity with the sequence of 503 bp at the 3'-end of the rhesus TPI gene (Old and Mohrenweiser, 1988).

Intriguingly, an apparent open reading frame (ORF) of 125 codons devoid of introns occurs 158 bp downstream (from +3424 to +3798 in Fig. 7) from the TPI gene AATAAA polyadenylation signal. No significant nucleotide or protein sequence homologies were found when comparing this supposed ORF with GenBank (Version 55.0) and NBRF/PIR. The ORF does not have an AATAAA sequence in the sequenced 3'-terminal region, presumably occurring further downstream. This presumptive ORF has an unusual promoter element motif. A canonical GCCAAT sequence (starting at +3015) associated with binding the transcription factor CTF/NF-1 (Mitchell and Tjian, 1989) occurs 409 bp prior to the ATG initiation codon. An activating transcription factor (ATF)-binding motif (ACGTCAC, starting at +3054) occurs 370 bp upstream from the ORF. This element is associated with several adenoviral promoter motifs and also occurs in various cellular promoter motifs, as in the genes for *c-fos*, somatostatin, α-gonadotropin, and vasoactive intestinal peptide (Lee *et al.*, 1987). The ORF also has another presumptive CCAAT element (CCCAT, starting at +3249) as well as two potential TATA promoter regions (TATATAA, starting at +3338, and TATAA, starting at +3389). The human TPI gene 3'-flanking region also has these possible promoter elements in essentially the same relative positions as occurs in the chimpanzee sequence. The corresponding rhesus sequence

would appear to have some of these putative elements, but there is a C to T transition in the rhesus sequence corresponding to the putative ATF motif (ACGTCAC to ATGTCAC), and the sequenced 3'-flanking portion of the rhesus TPI gene extends only slightly beyond the AATAAA polyadenylation signal. This ORF would encode a polypeptide of 124 amino acids, and its nucleotide sequence showed no similarity to known sequences in the data banks. Although there is no evidence that this ORF is expressed, predictive secondary structure and hydropathy calculations were done using the algorithms of Kyte and Doolittle (1982) and Garnier *et al.* (1978), respectively. These analyses indicate a putative hydrophobic N-terminal region and hydrophilic C-terminal section, a predominance of β-strands separated by β-turns or random structures, and an absence of α-helices. Thus, a compact protein of anti-parallel β-sheet structure is predicted.

Nucleotide Sequence Analysis of Two Chimpanzee TPI Retropseudogenes

As shown in Figs. 6 and 7, the nucleotide sequences of 3940-bp and 2231-bp segments, respectively, of the genomic clones λψchpTPI-A and λψchpTPI-B that hybridized to the human TPI cDNA probe were determined. Both λψchpTPI-A and λψchpTPI-B were found to encompass TPI processed

retropseudogenes, since they both possess the hallmark characteristics of this class of retrosequences, primarily since they basically resemble defective cDNA copies of the fully processed TPI mRNA (Vanin, 1985; Weiner *et. al.*, 1986; Hull and Will, 1989).

Similar to the structural attributes associated with other processed retropseudogenes, the TPI retrosequence in λψchpTPI-A (Figs. 6B and 7) lacks intervening sequences, has a 3'-terminal oligo(A) tract derived from the parental mRNA (starting at residue +3286 in Fig. 7), and has flanking 15-bp direct repeat sequences (5'-AAGAGAGCTATTGTA-3', starting at nucleotides -50 and +3297 in Fig. 7). In fact, the retrosequence in λψchpTPI-A is essentially a *de facto* TPI cDNA sequence that allowed the deduction of many structural attributes of the TPI transcription unit. As shown in Fig. 7, the retrosequence in λψchpTPI-A has multiple base substitutions, and a number of insertions or deletions that would preclude a functional TPI polypeptide, even if the retrosequence could be expressed without appropriate promoter elements.

A comparison of the sequence of λψchpTPI-A with DNA sequences in GenBank identified several interesting structural elements in the flanking regions of the pseudogene in λψchpTPI-A. An *Alu* family member was found in the 5'-flanking region from positions -381 to -86, with the

same relative polarity as the processed pseudogene. The *Alu* member is flanked by 14-bp direct repeats of 5'-AATACAAAATTTCT-3' (Fig. 7). It has been observed that the chimpanzee and human share some *Alu* repeats at identical sites in their genomes, and that their divergence rates are those expected for nonselected sequences (Sawada *et al.*, 1985; Koop *et al.*, 1986). Even though it has been estimated that there are approximately twice as many Alu repeats in human as in chimpanzee (Hwu *et al.*, 1986), both species share similar consensus sequences (Daniels *et al.*, 1983; Sawada *et al.*, 1985). The *Alu* member in λψchpTPI-A has 90% homology to the *Alu* consensus sequence derived by Quentin (1988), and can be classified into class E (Quentin, 1988), class II (Britten *et al.*, 1988), or class 5a (Jurka and Smith, 1988). These three classifications indicate an *Alu* member of intermediate age, originating approximately 35 million years ago (Labuda and Striker, 1989).

The sequence of λψchpTPI-A encompasses a region of DNA with significant similarity to a commonly occurring class of repetitive sequences variously called LINES (long interspersed elements), L1 elements or *KpnI* family members (Scott *et al.*, 1987; Fanning and Singer, 1987). Comparison of the chimpanzee sequence with that of one human *KpnI* family member (lg-kpni-7; DiGiovanni *et al.*, 1983) reveals a 66% identity from -762 to -673 (Fig. 7), a 58% identity from -641 to -396, and a 58% identity from +3398 to +3565.

The chimpanzee L1 sequence is interrupted between -677 and -632 by a stretch of 46 alternating purine-pyrimidine residues which has the potential to adopt the Z-DNA conformation, as well as by the DNA segment (flanked by the stretch of A residues centered at -398 and +3399) which encompasses both the *Alu* element and the processed pseudogene (which ends in the A-rich region centered around +3291). The L1 sequence preceding the *Alu* element is similar to that at the end of the second ORF (open reading frame) and the beginning of the 3'-untranslated region (3'-UTR) of a human L1 consensus sequence (Scott *et al.*, 1987), and the L1 sequence following the chimpanzee TPI pseudogene is similar to that of the remaining 3'-UTR. As is the case with about 95% of such L1 elements, the retroposon sequence is truncated at the 5'-side (Fanning and Singer, 1987) and no significant homology with other L1 sequences occurs prior to the -762 to -673 stretch. Regions of DNA sequence showing similarity to L1 elements have been previously found associated with Alu elements (Miyake *et al.*, 1983) as well as processed pseudogenes (Scarpulla, 1985).

Other regions of the DNA sequence of λψchpTPI-A were noted for their similarity to sequences in GenBank and/or the medium reiteration frequency repetitive elements called MER sequences (Jurka, 1990). The sequence between -973 and -871 has a 56% similarity to the sequence (21052 to 21155)

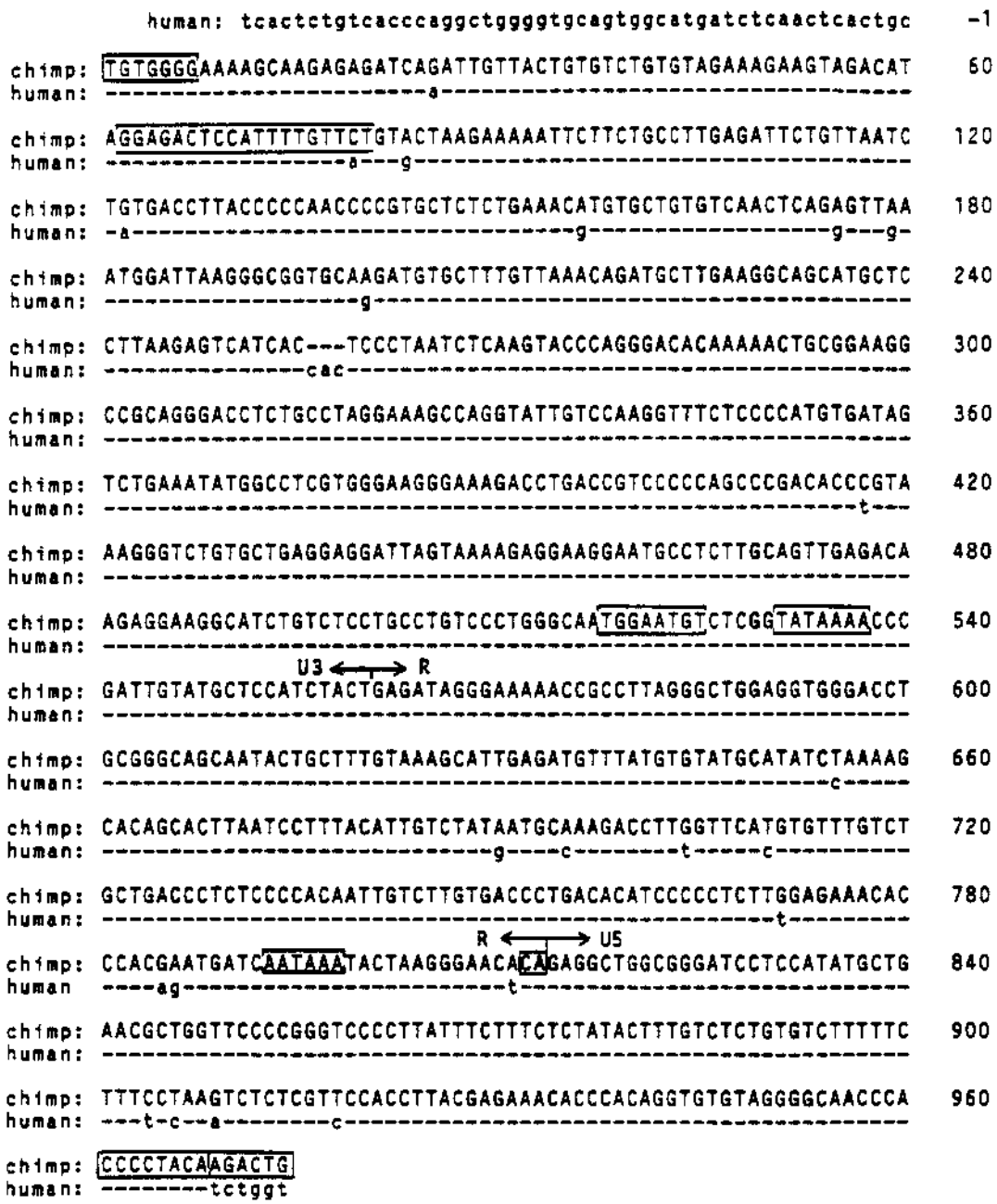
in an intron of the human tissue plasminogen activator gene (GenBank no. HUMTPA; Degen *et al.*, 1986), and thus has a significant similarity to MER4 repetitive sequences (Jurka, 1990). The complementary strand from -778 to -755 corresponds to a segment of MER3 repetitive sequences (Jurka, 1990). The DNA sequence from +3432 to +3525 has 81% identity with 94 bp at the 3'-end of the intron between the first and second hinges of the human IgD delta heavy chain gene (White *et al.*, 1985). Also, the region of +3693 to +3834 has a 68% similarity to a stretch of DNA sequence 247 bp upstream from a glycine $tRNA_{GCC}$ pseudogene (Pirtle *et al.*, 1986). The DNA sequence from +4019 to +4210 has 75% similarity to a region within an intron of the human apolipoprotein CIII gene (GenBank no. HUMAPC3G; Shelley *et al.*, 1985), with one stretch of 53 bp having 91% identity. Part of this region also has a significant similarity to the sequences of MER1 repetitive elements (Jurka 1990). The significance of these similarities is unclear, but it would appear that both sequenced flanks of the retrosequence in λψchpTPI-A consist primarily of a number of repetitive DNA sequences.

The TPI retrosequence in λψchpTPI-B (Figs. 6C and 7) also has many structural features associated with processed retropseudogenes, primarily that of having a defective but uninterrupted cDNA copy of the TPI mRNA. However, as

depicted in Fig. 6C, the 3'-terminal region of the chimpanzee genomic DNA harbored in λψchpTPI-B was ligated to the left arm of the λ Charon 32 vector during construction of the partial *EcoRI* genomic library. For this reason, the precise hallmark direct repeat sequences could not be inferred. The coding region of this TPI pseudogene lacks the last nine codons (from +2809 to +2835 in Fig. 7), and the 3'-flanking region with the characteristic oligo(A) tract is also missing. The defects in the coding region include typical base substitutions, deletions, or insertions, that would preclude a functional TPI polypeptide even if expressed.

Amazingly, a 980-bp insertion at +2212 (Fig. 7) corresponding to a solitary LTR (Fig. 8) was found to occur in the TPI coding portion of the retrosequence. A comparison of the 980-bp portion of the 2231-bp sequence of λψchpTPI-B with GenBank revealed identities of 97% with both the 5'- and 3'-LTRs of the human retrovirus gene HERV-K10 (Ono, 1986). As shown in Figs. 7 and 8, the LTR is flanked by direct repeats of 5'-AGACTG-3', including part of the region corresponding to Lys 174 - Ala 176. The ends of the LTR consist of imperfect inverted repeats, 5'-TGTGGGG...CCCCTACA-3' which possess the 5'-TG....CA-3' ends commonly found in eukaryotic LTRs (Temin, 1981; Varmus, 1982). Beginning at position 531, a TATA consensus sequence

Fig. 8. Comparison of the chimpanzee solitary long terminal repeat (LTR) in the λψchpTPI-B pseudogene with the LTR sequences from the human endogenous retrovirus gene HERV-K10 (Ono, 1986; Ono et al., 1986; GenBank accession number HUMERVKA). Dashes in the human LTR DNA sequences indicate a match with the chimpanzee LTR DNA sequence, and mismatched bases are given in lower-case letters, with the exception of residue 296 in the human 3'-LTR sequence, which is A instead of G. The cellular direct repeat sequence (AGACTG) is boxed only at the LTR 3'-terminus, since the 5'-terminal direct repeat is a part of the defective coding region in λψchpTPI-B shown in Fig. 2. The LTR U3, R, and U5 segments are indicated by arrows (Temin, 1981; Chen and Barker, 1984; Ono, 1986). The imperfect inverted terminal repeats TGTGGGG and CCCCTACA start at 1 and 961, respectively. Other interesting features (boxed) include: a variant glucocorticoid response element, GGAGACTCCATTTTGTTCT, starting at residue 62 (Yamamoto, 1985), a SV40 core enhancer element, TGGAATGT, starting at 518 (Khoury and Gruss, 1983), a TATA promoter element (TATAAAA, starting at 531), a presumptive polyadenylation signal (AATAAA, starting at 793), and a putative polyadenylation site (CA, residues 812-813).

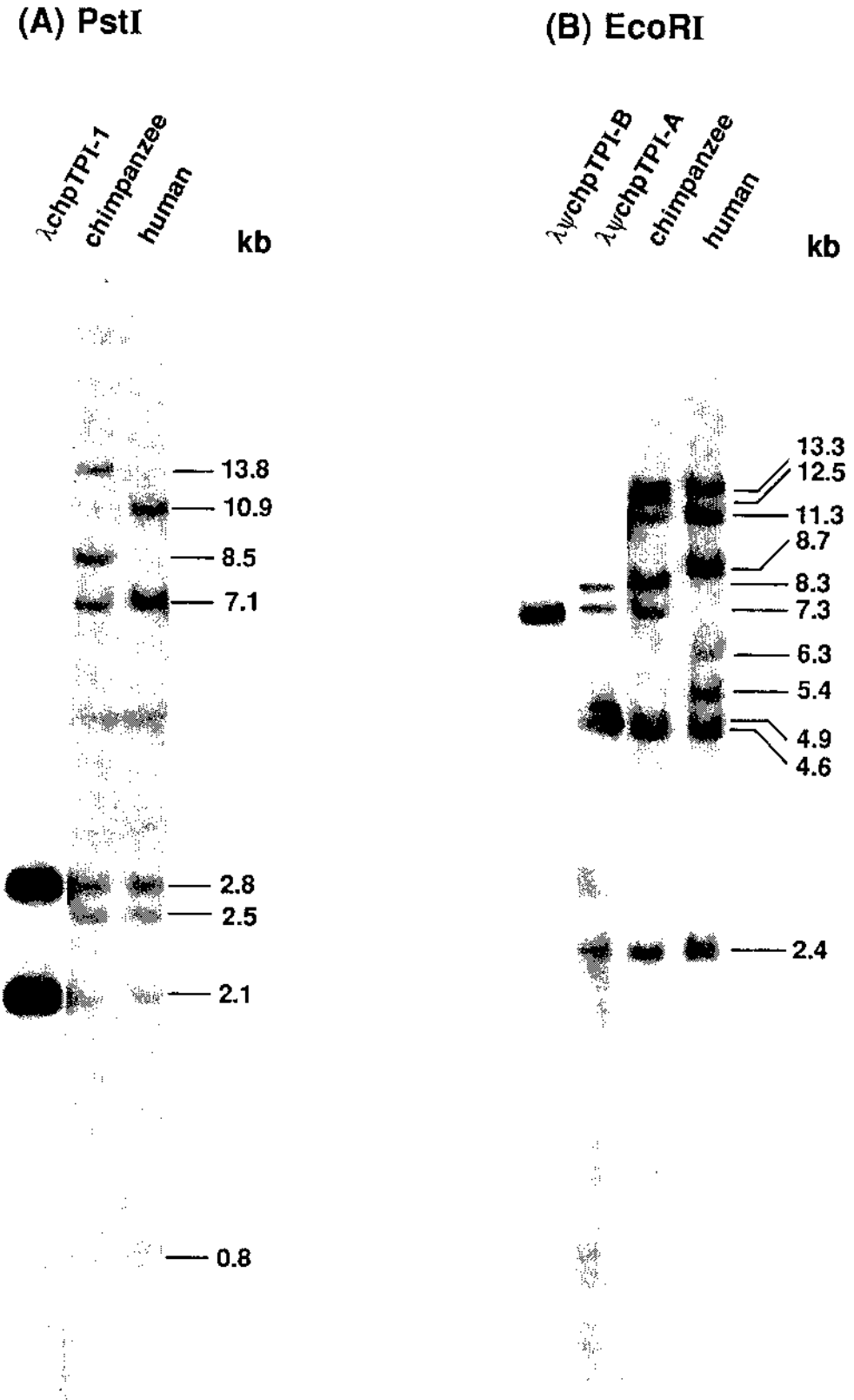

human: tcactctgtcacccaggctggggtgcagtggcatgatctcaactcactgc -1
chimp: TGTGGGGAAAAGCAAGAGAGATCAGATTGTTACTGTGTCTGTGTAGAAAGAAGTAGACAT 60
human: ----------a----------
chimp: AGGAGACTCCATTTTGTTCTGTACTAAGAAAAATTCTTCTGCCTTGAGATTCTGTTAATC 120
human: ----------a---g----------
chimp: TGTGACCTTACCCCCAACCCCGTGCTCTCTGAAACATGTGCTGTGTCAACTCAGAGTTAA 180
human: -a----------g----------g---g-
chimp: ATGGATTAAGGGCGGTGCAAGATGTGCTTTGTTAAACAGATGCTTGAAGGCAGCATGCTC 240
human: ----------g----------
chimp: CTTAAGAGTCATCAC---TCCCTAATCTCAAGTACCCAGGGACACAAAAACTGCGGAAGG 300
human: ----------cac----------
chimp: CCGCAGGGACCTCTGCCTAGGAAAGCCAGGTATTGTCCAAGGTTTCTCCCCATGTGATAG 360
human: ----------
chimp: TCTGAAATATGGCCTCGTGGGAAGGGAAAGACCTGACCGTCCCCCAGCCCGACACCCGTA 420
human: ----------t---
chimp: AAGGGTCTGTGCTGAGGAGGATTAGTAAAAGAGGAAGGAATGCCTCTTGCAGTTGAGACA 480
human: ----------
chimp: AGAGGAAGGCATCTGTCTCCTGCCTGTCCCTGGGCAATGGAATGTCTCGGTATAAAACCC 540
human: ----------
U3 ←→ R
chimp: GATTGTATGCTCCATCTACTGAGATAGGGAAAAACCGCCTTAGGGCTGGAGGTGGGACCT 600
human: ----------
chimp: GCGGGCAGCAATACTGCTTTGTAAAGCATTGAGATGTTTATGTGTATGCATATCTAAAAG 660
human: ----------c-----
chimp: CACAGCACTTAATCCTTTACATTGTCTATAATGCAAAGACCTTGGTTCATGTGTTTGTCT 720
human: ----------g----c--------t-----c----------
chimp: GCTGACCCTCTCCCCACAATTGTCTTGTGACCCTGACACATCCCCCTCTTGGAGAAACAC 780
human: ----------t---------
R ←→ U5
chimp: CCACGAATGATCAATAAATACTAAGGGAACACAGAGGCTGGCGGGATCCTCCATATGCTG 840
human ----ag----------t----------
chimp: AACGCTGGTTCCCCGGGTCCCCTTATTTCTTTCTCTATACTTTGTCTCTGTGTCTTTTTC 900
human: ----------
chimp: TTTCCTAAGTCTCTCGTTCCACCTTACGAGAAACACCCACAGGTGTGTAGGGGCAACCCA 960
human: ---t-c--a--------c----------
chimp: CCCCTACAAGACTG
human: --------tctggt

is found, but no CCAAT element occurs 40 to 50 bp upstream from the TATA motif in this LTR. A polyadenylation signal, AATAAA, is found starting at position 793, and the presumed polyadenylation site, CA, is located 13 bp downstream from the polyadenylation signal. Other structural elements include a variant glucocorticoid response element, 5'-GGAGACTCCATTTTGTTCT-3', starting at residue 62 (Yamamoto, 1985), and a SV40 core enhancer element, 5'-TGGAATGT-3', starting at 518 (Khoury and Gruss, 1983). Solitary LTRs have been previously observed (Hughes *et al.*, 1981; Kuemmerle *et al.*, 1987), and are thought to be the result of excision of the provirus from the genome by homologous recombination within the LTRs.

Southern Blot Hybridization Analysis of Genomic and Cloned DNAs

Southern blot hybridization analysis was done in order to relate the sizes of the cloned chimpanzee DNA fragments with the potential corresponding chimpanzee and human genomic DNA fragments, and to estimate the number of members in the chimpanzee TPI gene family. As shown in Fig. 9A, *PstI* digestion produces six genomic DNA fragments, from both chimpanzee and human, that hybridize to the human TPI cDNA probe. Four of the hybridizing fragments apparently are identical in size (2.1-, 2.5-, 2.8-, and 7.1-kb) in the two species. In contrast, chimpanzee DNA has two unique

Fig. 9. Comparison of cloned chimpanzee TPI sequences with chimpanzee and human genomic DNAs by Southern blot hybridization. The recombinant chimpanzee DNAs, harbored in λchpTPI-1, λψchpTPI-A, and λψchpTPI-B (0.2 μg), and chimpanzee and human genomic DNAs (10 μg) were digested with (A) *PstI* or (B) *EcoRI*. The restricted DNAs were fractionated on 0.8% agarose gels, transferred to positively-charged nylon membranes, and hybridized with ^{32}P-labeled DNA fragments derived from an upstream segment of the human TPI cDNA clone pHTPI-5a (Maquat *et al.*, 1985). The sizes (in kilobases) of the fragments shown were determined relative to standard markers generated by digesting λ DNA with *HindIII* and pBR322 DNA with *HinfI*.


(A) PstI
λchpTPI-1
chimpanzee
human
kb
13.8
10.9
8.5
7.1
2.8
2.5
2.1
0.8
(B) EcoRI
λψchpTPI-B
λψchpTPI-A
chimpanzee
human
kb
13.3
12.5
11.3
8.7
8.3
7.3
6.3
5.4
4.9
4.6
2.4

fragments (8.5- and 13.8-kb) which are not present in human DNA, whereas human DNA also has two fragments (0.8- and 10.9-kb) which are absent in chimpanzee DNA. The hybridizing 2.1- and 2.8-kb *PstI* chimpanzee and human genomic fragments can be seen to correlate exactly with the fragments generated by *PstI* digestion of the chimpanzee genomic clone λchpTPI-1 encompassing the TPI transcription unit, and are similar in size to the homologous fragments observed by Maquat *et al.* (1985) for the human TPI genomic clone λhTPI-8B. Furthermore, the sizes of these two *PstI* fragments correlate very well to the actual sizes as determined by DNA sequence analysis (1975-bp and 2767-bp, respectively). No *PstI* restriction site occurs in the coding region of the chimpanzee functional gene, although there is a *PstI* site in intron 4 (Figs. 6, 7). Since the processed pseudogene in λψchpTPI-B is adjacent to the left λ Charon 32 arm, *PstI* digestion of λψchpTPI-B DNA generates a 5.3-kb hybridizing fragment which contains a 2.1-kb segment derived from the left λ Charon 32 arm (Loenen and Blattner, 1983), and does not directly correspond to a hybridizing genomic fragment. *PstI* digestion of λψchpTPI-A DNA generates a hybridizing 10.5-kb fragment (not shown) which also does not have a corresponding hybridizing genomic fragment, presumably due to the presence of a segment of the left vector arm in the fragment. Thus, neither of the chimpanzee TPI pseudogenes

can be associated with a *PstI*-generated genomic fragment. It should also be mentioned that the hybridizing 2.5-kb *PstI* fragments derived from the chimpanzee and human genomic DNAs very likely correspond to a similar-sized *PstI* fragment observed by Maquat *et al.* (1985) that contains at least part of the human TPI pseudogene ψhTPI-13C.

EcoRI digestion generates approximately seven chimpanzee and six human genomic DNA fragments that hybridize to the probe derived from the human TPI cDNA, as shown in Fig. 9B. Three of the hybridizing fragments from each of the two species appear to be identical in size (2.4- , 4.6-, and 11.3-kb). *EcoRI* digestion of λψchpTPI-B DNA precisely removes the left λ arm (Fig. 6C) and produces a 7.3-kb hybridizing fragment which directly corresponds to a hybridizing fragment in the chimpanzee genomic DNA digest. Digestion of λψchpTPI-A DNA with *EcoRI* produces a strongly hybridizing 4.9-kb fragment that clearly corresponds to a chimpanzee genomic fragment. A less intense 2.4-kb hybridizing fragment was also observed in the *EcoRI* digest of λψchpTPI-A DNA, potentially corresponding to the 2.4-kb hybridizing fragments in chimpanzee and human genomic DNAs (Fig. 9B). However, this light band is thought to be a contaminant, since DNA sequence analysis of this segment of λψchpTPI-A indicated no *EcoRI* sites within the region corresponding to the probe, and since other restriction

endonuclease digests of λψchpTPI-A DNA indicated only one region homologous to the TPI cDNA probe. Also, the two hybridizing fragments about 8-kb in size in the λψchpTPI-A lane of Fig. 9B are also thought to be contaminants, since these bands are also light on the photograph of the stained agarose gel.

Maquat *et al.* (1985) deduced that there is a single functional human TPI gene (hTPI-8B), and that it occurs within 10.0- and 1.6-kb *EcoRI* fragments, that one human TPI pseudogene (ψhTPI-5A) is encompassed within 5.7- and 4.3-kb *EcoRI* fragments, and that another human TPI processed pseudogene (ψhTPI-19A) is contained within a 2.3-kb *EcoRI* fragment. The hybridizing 2.4-kb *EcoRI* fragment (Fig. 9B) in the chimpanzee and human genomic DNAs very likely is homologous to the similar-sized band observed by Maquat *et al.* (1985) for the TPI pseudogene ψhTPI-19A, whereas the hybridizing 6.3-kb and 4.6-kb *EcoRI* human genomic fragments in Fig. 9B correspond to similar-sized *EcoRI* fragments detected by Maquat and coworkers (1985) for human TPI pseudogene ψhTPI-5A. It is our conjecture that the 7.3-kb *EcoRI* fragment observed in chimpanzee genomic DNA and in λψchpTPI-B is related to the human 6.3-kb *EcoRI* fragment. This can be attributed to the presence of the LTR sequence in λψchpTPI-B (Fig. 8), and, thus, the hybridizing fragment in chimpanzee DNA would be anticipated to be about 1.0-kb

larger than the corresponding human DNA fragment. Digestion of λchpTPI-1 DNA with *EcoRI* generates two very intensely hybridizing fragments (13.3-kb and 11.3-kb) with the human TPI cDNA probe. A less intensely hybridizing 1.7-kb band was also observed in the same autoradiogram. It would appear that the 11.3-kb and 1.7-kb *EcoRI* fragments derived from λchpTPI-1 correspond to the similar-sized fragments observed by Maquat *et al.* (1985) for the single human TPI gene. The 13.3-kb hybridizing fragment observed in Fig. 9B can be accounted for by being a partial product containing both the 11.3-kb and 1.7-kb hybridizing fragments. This could also explain the absence of the 1.7-kb band in the light exposure of the autoradiogram depicted in Fig. 9B. Genomic blotting experiments with chimpanzee and human DNAs, using *EcoRI* and a probe derived from intron 1 of the sequenced chimpanzee TPI gene, indicated hybridizing fragments of 13.3 kb and 11.3 kb, the 13.3-kb band presumably being a partial product. Thus, from the above analyses, we conclude that there are six or seven homologous TPI loci in the chimpanzee genome and a similar, if not identical, number in the human genome, and that a single functional TPI gene also occurs in the chimpanzee genome. This is in excellent agreement with the genomic blotting results of Maquat *et al.* (1985).

Evolution of the TPI Retropseudogenes in λψchpTPI-A and λψchpTPI-B

The percent divergence calculated between the TPI pseudogene in λψchpTPI-A and the chimpanzee TPI coding region was determined to be 1.8%, corresponding to an approximate age of 2.6 million years. Since the divergence of chimpanzee and human is believed to have occurred between 5.5 and 7.7 million years ago (Lewin, 1988), these estimates indicate that this retrosequence was integrated into the genome after the split of the chimpanzee and human lineages. This prediction is consistent with the genomic blotting results (Fig. 9B), since the retrosequence in λψchpTPI-A is contained within a 4.9-kb chimpanzee genomic *EcoRI* fragment, which does not have a corresponding human genomic DNA fragment.

The percent divergence calculated for λψchpTPI-B was 7.3%, corresponding to an approximate age of 10.4 million years. This estimate suggests that the pseudogene in λψchpTPI-B integrated into the genome prior to the divergence of chimpanzee and human. Since the flanking region of the human TPI processed pseudogene ψhTPI-5A (Brown *et al.*, 1985) has a 98% similarity to that of λψchpTPI-B, it can be conjectured that these two processed pseudogenes have a common ancestral element and are the product of a single

retroposition event. The genomic blotting results are consistent with these conclusions. The TPI retrosequence in λψchpTPI-B was shown to be contained within a 7.3-kb chimpanzee genomic *EcoRI* fragment (Fig. 9B). Since the LTR found within λψchpTPI-B is not present in the human TPI pseudogene ψhTPI-5A, its insertion into the chimpanzee genome was apparently due to events following the divergence of the two species. The human TPI pseudogene ψhTPI-5A is encompassed within an *EcoRI* fragment in the human genome, which corresponds to the human 6.3-kb band in Fig. 9B. Due to the presence of the LTR in λψchpTPI-B, the hybridizing chimpanzee fragment is expected to be approximately 1.0 kb longer than the corresponding human fragment.

Since the TPI pseudogenes in λψchpTPI-B and λψhTPI-5A apparently have a common ancestor which was integrated into the genome prior to the divergence of chimpanzee and human, all differences between the two processed pseudogenes have accumulated following the divergence of the two species. A comparison of the nucleotide sequences of λψchpTPI-A and ψhTPI-5A, both of which are assumed to be changing in a nonselected manner, shows a divergence of 2.2%. Nucleotide changes per site were corrected for multiple events by the formula $d = -3/4 \ln[1-(4/3)p]$, where p is the uncorrected value and d is the corrected value (Jukes and Cantor, 1969). This value is in agreement with the expected divergence for

nonselected human and chimpanzee DNAs (Chang and Slightom, 1984; Willard *et al.*, 1985).

The process of retroposition appears to be a relatively recent development on the evolutionary time scale. The reason for fixation of retropseudogenes in eukaryotic genomes is uncertain. The most likely explanation is that the resulting variations due to retroposition provides for evolutionary change by creating new material for positive or negative selection, as well as neutral drift. Processed pseudogenes, with few exceptions, have been found only in mammals, and appear to be a common phenomenon with housekeeping genes which exhibit germ-line expression. SINEs and LINEs, which also appear to be the result of retroposition, have apparently developed concurrently. The *Alu* family member, the L1 element, the MER elements, the solitary LTR, and the two processed pseudogenes described above are the products of different retroposition events in the chimpanzee genome. Even though many of the details of retroposition are poorly understood at this time, it is apparent that numerous retroposition events have occurred in these two different segments of the chimpanzee genome, and undoubtedly were the significant events in the divergence of these two genomic regions.

The Elusive Origin of the Hominoid TPI-2 Polypeptide

Two basic genetic models have previously been proposed

for the expression of the two hominoid isozymes of TPI. One model suggests that TPI-1 and TPI-2 are the products of two distinct structural loci (Kester *et al.*, 1977; Yuan *et al.*, 1979). The other model intimates that both TPI-1 and TPI-2 are the products of a single structural locus (Decker and Mohrenweiser, 1981; Asakawa and Iida, 1985). The primary evidence for a single structural gene is based on studies with electrophoretic variants of TPI. The variant phenotype of the constitutive TPI isozyme is also found in the proliferation-specific isozyme. In order for the two-loci hypothesis to be consistent with this result, both loci must have the variant form. It could be argued that if two structural loci do exist, they could be correcting each other by a process such as gene conversion, which could result in both loci maintaining identical phenotypes.

It has been suggested that the proliferation specific TPI-2 isozyme could be due to an alternative splicing mechanism (Decker and Mohrenweiser, 1986), but the DNA sequences of the characterized primate TPI genes do not apparently have any presumptive alternative splice junction sites which would support this hypothesis. In addition, Old *et al.* (1989) observed only a single TPI mRNA species, in a variety of human tissue culture cells and African green monkey kidney fibroblast cells, similar in size to the low-abundance, 1250-nucleotide human TPI mRNA detected by Brown *et al.* (1985). Thus, Old *et al.* (1989) discounted a variety

of possible mechanisms for synthesis of TPI-2 in proliferating hominoid cells, such as differential splicing of the primary TPI mRNA transcript and co-translational modification of the TPI-1 to the TPI-2 subunit, and commented that either the primary transcript of the structural locus or the gene product must be the substrate for the modification of TPI-1 to TPI-2. If only one structural locus exists, and, if differential splicing is discounted, then a mechanism could be conjectured that would involve a locus linked with the TPI structural gene that would in some manner encode a modifier RNA or polypeptide. Both RNA synthesis and *de novo* protein synthesis are required for TPI-2 expression (Kester *et al.*, 1977). The presence of a presumptive ORF immediately downstream from the TPI gene at the same chromosomal location is consistent with the previous observations, but it would be necessary to invoke the mechanism involving a linked modification RNA or polypeptide. Thus, it is interesting to speculate that the potential ORF in the 3'-flanking region of the chimpanzee TPI gene, if expressed, could encode a product somehow involved in the production of the proliferation-specific TPI-2 subunit observed in the hominoids, and that the putative ATF element in the corresponding rhesus sequence could be rendered nonfunctional due to a single base substitution, possibly leading to the nonexpression of TPI-2 in nonhominoids expressing only TPI-1. Therefore, at this

time, the precise mechanism responsible for the appearance of TPI-2 in proliferating hominoid cells still remains an enigma.

BIBLIOGRAPHY

ALBER, T., and KAWASAKI, G. (1982). Nucleotide sequence of the triose phosphate isomerase gene of *Saccharomyces cerevisiae*. J. Molec. Appl. Genet. **1**, 419-434.

ALBERY, W.J., and KNOWLES, J.R. (1977). Perfection in Enzyme Catalysis: The energetics of triosephosphate isomerase. Accts. Chem. Res. **10**, 105-111.

ARNOLD, G.J., and GROSS, H.J. (1987). Unrelated leader sequences can efficiently promote human tRNA gene transcription. Gene **51**, 237-246.

ARNOLD, G.J., SCHMUTZLER, C., and GROSS, H.J. (1988). Extragenic control regions of human $tRNA^{Val}$ genes reveals two different regulatory effects. DNA **7**, 87-97.

ARNOLD, G.J., SCHMUTZLER, C., THOMANN, U., VAN TOL, H., and GROSS, H.J. (1986). The human $tRNA^{Val}$ gene family: organization, nucleotide sequences and homologous transcription of three single-copy genes. Gene **44**, 287-297.

ARTAVANIS-TSAKONAS, S., and HARRIS, J.I. (1980). Primary structure of triosephosphate isomerase from *Bacillus stearothermophilus*. Eur. J. Biochem. **108**, 599-611.

ASAKAWA, J., and IIDA, S. (1985). Origin of human triosephosphate isomerase isozymes: further evidence for the single structural locus hypothesis with Japanese variants. Hum. Genet. **71**, 22-26.

BARNES, W.M., BEVAN, M., and SON, P.H. (1983). Kilo-sequencing: Creation of an ordered nest of asymmetric

deletions across a large target sequence carried on phage M13. Methods Enzymol. **101**, 98-122.

BARRELL, B.G., and SANGER, F. (1969). The sequence of phenylalanine tRNA from *E. coli*. FEBS Lett. **3**, 275-278.

BENTON, W.D., and DAVIS, R.W. (1977). Screening λgt recombinant clones by hybridization to single plaques *in situ*. Science **196**, 180-182.

BIRNBOIM, H.C., and DOTY, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucl. Acids Res. **7**, 1513-1523.

BLATTER, F.R., BLECHL, A.E., DENNISTON-THOMPSON, K.,FABER, H.E., RICHARDS, J.E., SLIGHTOM, J.L., TUCKER, P.W., and SMITHIES, O. (1978). Cloning human fetal γ-globin and mouse α-type globin DNA: preparation and screening of shotgun collections. Science **202**, 1279-1284.

BLIN, N., and STAFFORD, D.W. (1976). A general method for isolation of high molecular weight DNA from eukaryotes. Nucl. Acids Res. **3**, 2303-2308.

BOYER, T.G., KRUG, J.R., and MAQUAT, L.E. (1989). Transcriptional regulatory sequences of the housekeeping gene for human triosephosphate isomerase. J. Biol. Chem. **264**, 5177-5187.

BRITTEN, R.J., BARON, W.F., STOUT, D.B., and DAVIDSON, E. H. (1988). Sources and evolution of human *Alu* repeated sequences. Proc. Natl. Acad. Sci. USA **85**, 4770-4774.

BROWN, J.R., DAAR, I.O., KRUG, J.R., and MAQUAT, L.E. (1985). Characterization of the functional gene and several processed pseudogenes in the human triosephosphate isomerase gene family. Mol. Cell. Biol. **5**, 1694-1706.

CAPONE, J. P. (1988). Modulation of the phenotypic

expression of a human serine tRNA gene by 5'-flanking sequences. DNA **7**, 459-468.

CHANG, L.-Y.E., and SLIGHTOM, J.L. (1984). Isolation and nucleotide sequence analysis of the β-type globin pseudogene from human, gorilla and chimpanzee. J. Mol. Biol. **180**, 767-784.

CHANG, Y.N., PIRTLE, I.L., and PIRTLE, R.M. (1986). Nucleotide sequence and transcription of a human tRNA gene cluster with four genes. Gene **48**, 165-174.

CHEN, E.Y., and ROE, B.A. (1977). Sequence studies on human placenta $tRNA^{Val}$: Comparison with the mouse myeloma $tRNA^{Val}$. Biochem. Biophys. Res. Commun. **78**, 631-640.

CHEN, E.Y., and SEEBURG, P.H. (1985). Supercoil sequencing: A fast and simple method for sequencing plasmid DNA. DNA **4**, 165-170.

CHEN, H.R., and BARKER, W.C. (1984). Nucleotide sequences of the retroviral long terminal repeats and their adjacent regions. Nucl. Acids Res. **12**, 1767-1778.

CORRAN, P.H., and WALEY, S.G. (1975). The amino acid sequence of rabbit muscle triose phosphate isomerase. Biochem. J. **145**, 335-344.

CRAIG, L.C., WANG, L.P., LEE, M.M., PIRTLE, I.L., and PIRTLE, R.M. (1989). A human tRNA gene cluster encoding the major and minor valine tRNAs and a lysine tRNA. DNA **8**, 457-471.

DALE, R.M.K., McCLURE, B.A., and HOUCHINS, J.P. (1985). A rapid single-stranded cloning strategy for producing a sequential series of overlapping clones for use in DNA sequencing: application to sequencing the corn mitochondrial 18S rDNA. Plasmid **13**, 31-40.

DANIELS, G.R., and DEININGER, P.L. (1985). Integration site preferences of the *Alu* family and similar repetitive DNA sequences. Nucl. Acids Res. **13**, 8939-8954.

DANIELS, G.R., FOX, M., LOWENSTEINER, D., SCHMID, C., and DEININGER, P.L. (1983). Species-specific homogeneity of the primate *Alu* family of repeated DNA sequences. Nucl. Acids Res. **11**, 7579-7593.

DAS, G., HENNING, D., WRIGHT, D., and REDDY, R. (1988). Upstream regulatory elements are necessary and sufficient for transcription of a U6 RNA gene by RNA polymerase III. EMBO J. **7**, 503-512.

DAYHOFF, M.O. (1978). Atlas of Protein Sequence and Structure, Vol. 5, Supplement 3. Georgetown Univ. Med. Center, Washington, D.C.

DECKER, R.S., and MOHRENWEISER, H.W. (1981). Origin of the triosephosphate isomerase isozymes in humans: genetic evidence for the expression of a single structural locus. Am. J. Hum. Genet. **33**, 683-691.

DECKER, R.S., and MOHRENWEISER, H.W. (1986). Hominoid triosephosphate isomerase: Characterization of the major cell proliferation specific isozyme. Mol. Cell. Biochem. **71**, 31-44.

DEGEN, S.J.F., RAJPUT, B., and REICH, E. (1986). The human tissue plasminogen activator gene. J. Biol. Chem. **261**, 6972-6985.

DE WET, J.R., DANIELS, D.L., SCHROEDER, J.L., WILLIAMS, B.G., DENNISTON-THOMPSON, K., MOORE, D.D., and BLATTNER, F.R. (1980). Restriction maps for twenty-one Charon vector phages. J. Virology **33**, 401-410.

DIGIOVANNI, L., HAYNES, S.R., MISRA, R., and JELINEK, W.R. (1983). *KpnI* family of long-dispersed repeated DNA

sequences of man: evidence for entry into genomic DNA of DNA copies of poly(A)-terminated *KpnI* RNAs. Proc. Natl. Acad. Sci. USA **80**, 6533-6537.

DORAN, J.L., WEI, X., and ROY, K.L. (1987). Analysis of a human gene cluster coding for $tRNA^{Phe}_{GAA}$ and $tRNA^{Lys}_{UUU}$. Gene **56**, 231-243.

DYNAN, W.S. (1986). Promoters for housekeeping genes. Trends Genet. **2**, 196-197.

EFSTRATIADIS, A., POSAKONY, J.W., MANIATIS, T., LAWN, R.M., O'CONNELL, C., SPIRITZ, R.A., DERIEL, J.K., FORGET, B.G., SLIGHTON, L., BLECHL, A.E., SMITHIES, O., BARALLE, F.E., SHOULDERS, C.C., and PROUDFOOT, N.J. (1980). The structure and evolution of the human β-globin gene family. Cell **21**, 653-668.

ELDER, J.T., PAN, J., DUNCAN, C.H., and WEISSMAN, S.M. (1981). Transcriptional analysis of interspersed repetitive polymerase III transcription units in human DNA. Nucl. Acids Res. **9**, 1171-1189.

ELLIOTT, M.S., and TREWYN, R.W. (1984). Inosine biosynthesis in transfer RNA by an enzymatic insertion of hypoxanthine. J. Biol. Chem. **259**, 2407-2410.

FANNING, T.G., and SINGER, M.F. (1987). LINE-1: a mammalian transposable element. Biochim. Biophys. Acta **910**, 203-212.

FEINBERG, A.P., and VOGELSTEIN, B. (1983). A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **132**, 6-13.

FURTH, A.J., MILMAN, J.D., PRIDDLE, J.D., and OFFORD, R.E. (1974). Studies on the subunit structure and amino acid sequence of triose phosphate isomerase from chicken breast muscle. Biochem. J. **139**, 11-25.

GARNIER, J., OSGUTHORPE, D.J., and ROBSON, B. (1978). Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. J. Mol. Biol. **120**, 97-120.

GEIDUSCHEK, E.P., and TOCCHINI-VALENTINI, G.P. (1988). Transcription by RNA polymerase III. Annu. Rev. Biochem. **57**, 873-914.

GODSON, G.N., and VAPNEK, D. (1973). A simple method of preparing large amounts of ϕX174RFI supercoiled DNA. Biochem. Biophys. Acta **299**, 516-520.

GRAUR, D., SHUALI, Y., and LI, W.-H. (1989). Deletions in processed pseudogenes accumulate faster in rodents than in humans. J. Mol. Evol. **28**, 279-285.

GRUNSTEIN, M., and HOGNESS, D.S. (1975). Colony hybridization: A method for the isolation of cloned DNAs that contain a specific gene. Proc. Natl. Acad. Sci. USA **72**, 3961-3965.

HANAHAN, D.J. (1983). Studies on transformation of *Escherichia coli* with plasmids. J. Mol. Biol. **166**, 557-580.

HATLEN, L., and ATTARDI, G. (1971). Proportion of the HeLa cell genome complementary to transfer RNA and 5S RNA. J. Mol. Biol. **56**, 535-553.

HEDGCOTH, C., HAYENGA, K., HARRISON, M., and ORTWERTH, B.J. (1984). Lysine tRNAs from rat liver: Lysine tRNA sequences are highly conserved. Nucl. Acids Res. **12**, 2535-2541.

HENIKOFF, S. (1984). Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene **28**, 351-359.

HUGHES, S.H., TOYOSHIMA, K., BISHOP, J.M., and VARNUS, H.E. (1981). Organization of the endogenous proviruses of chickens: implications for origin and expression. Virology **108**, 189-207.

HULL, R., and WILL, H. (1989). Molecular biology of viral and nonviral retroelements. Trends Genet. **5**, 357-359.

HWU, H.R., ROBERTS, J.W., DAVIDSON, E.H., and BRITTEN, R.J. (1986). Insertion and/or deletion of many repeated DNA sequences in human and higher ape evolution. Proc. Natl. Acad. Sci. USA **83**, 3875-3879.

JANK, P., RIESNER, D., and GROSS, H.J. (1977b). Rabbit liver $tRNA_1^{Val}$: II. Unusual secondary structure of TΨC stem and loop due to a U_{54}:A_{60} base pair. Nucl. Acids Res. **4**, 2009-2020.

JANK, P., SHINDO-OKADA, N., NISHIMURA, S., and GROSS, H.J. (1977a). Rabbit liver $tRNA_1^{Val}$: I. Primary structure and unusual codon recognition. Nucl. Acids Res. **4**, 1999-2008.

JOHNSON, G.D., PIRTLE, I.L., and PIRTLE, R.M. (1985). The nucleotide sequence of tyrosine $tRNA_{Q^*\Psi A}$ from bovine liver. Arch. Biochem. Biophys. **236**, 448-453.

JUKES, T.H., and CANTOR, C.R. (1969). Evolution of protein molecules. In: Munro, H.N. (ed.) Mammalian protein metabolism, Vol. III. Academic Press, New York, 21-123.

JURKA, J. (1990). Novel families of interpersed repetitive elements from the human genome. Nucl. Acids. Res. **18**, 137-141.

JURKA, J., and SMITH, T. (1988). A fundamental division in the *Alu* family of repeated sequences. Proc. Natl. Acad. Sci. USA **85**, 4775-4778.

KADONAGA, J.T., JONES, K.A., and TJIAN, R. (1986). Promoter-

specific activation of RNA polymerase II transcription by Sp1. Trends Biochem. Sci. **11**, 20-23.

KAHNT, B., FRANK, R., BLOCKER, H., and GROSS, H.J. (1989). An efficiently transciribed human $tRNA^{Val}$ gene variant produces a stable pre-tRNA: Repair of the processing defect by *in vitro* mutations. DNA **8**, 51-58.

KESTER, M.V., JACOBSON, E.L., and GRACY, R.W. (1977). The synthesis of a labile triosephosphate isomerase isozyme in human lymphoblasts and fibroblasts. Arch. Biochem. Biophys. **180**, 562-569.

KHOURY, G., and GRUSS, P. (1983). Enhancer elements. Cell **33**, 313-314.

KOLB, E., HARRIS, J.I., and BRIDGEN, J. (1974). Triosephosphate isomerase from coelacanth. Biochem. J. **137**, 185-197.

KOOP, B.F., MIYAMOTO, M.M., EMBURY, J.E., GOODMAN, M., CZELUSNIAK, J., and SLIGHTOM, J.L. (1986). Nucleotide sequence and evolution of the orangutan ε globin gene region and surrounding *Alu* repeats. J. Mol. Evol. **24**, 94-102.

KUEMMERLE, N.B., CH'ANG, L.-Y., KOH, C.K., BOONE, L.R., and YANG, W.K. (1987). Characterization of two solitary long terminal repeats of murine leukemia virus type that are conserved in the chromosome of laboratory inbred mouse strains. Virology **160**, 379-388.

KYTE, J., and DOOLITTLE, R.F. (1982). A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157**, 105-132.

LABUDA, D., and STRIKER, G. (1989). Sequence conservation in *Alu* evolution. Nucl. Acids Res. **17**, 2477-2491.

LARSON, R., and MESSING, J. (1983). Apple II computer software for DNA and protein sequence data. DNA **2**, 31-35.

LASKI, F.A., FIRE, A.Z., RAJBHANDARY, U.L., and SHARP, P. A. (1983). Characterization of tRNA precursor splicing in mammalian extracts. J. Biol. Chem. **258**, 11974-11980.

LEE, B.J., DE LA PENA, P., TOBIAN, J.A., ZASLOFF, M., and HATFIELD, D. (1987). Unique pathway of expression of an opal suppressor phosphoserine tRNA. Proc. Natl. Acad. Sci. USA **84**, 6384-6388.

LEE, K.A.W., HAI, T.-Y., SIVARAMAN, L., THIMMAPPAYA, B., HURST, H.C., JONES, N.C., and GREEN, M.R. (1987). A cellular protein, activating transcription factor, activates transcription of multiple E1A-inducible adenovirus early promoters. Proc. Natl. Acad. Sci. USA **84**, 8355-8359.

LEE, M.G.-S., LEWIS, S.A., WILDE, C.D., and COWAN, N.J. (1983). Evolutionary history of a multigene family: an expressed human β-tubulin gene and three processed pseudogenes. Cell **33**, 477-487.

LEGERSKI, R.J., HODNETT, J.L., and GRAY, H.B., JR. (1978). Extracellular nucleases of *Pseudomonas* BAL31. III. Use of the double-stranded deoxyriboexonuclease activity as the basis of a convenient method for the mapping of fragments of DNA produced by cleavage with restriction enzymes. Nucl. Acids Res. **5**, 1445-1464.

LEWIN, R.(1988). Molecular clocks turn a quarter century. Science **239**, 561-563.

LI, W.-H., GOJOBORI, T., and NEI, M. (1981). Pseudogenes as a paradigm of neutral evolution. Nature **292**, 237-239.

LIN, V.K., and AGRIS, P.F. (1980). Alterations in tRNA

isoaccepting species during erythroid differentiation of the Friend Leukemia cell. Nucl. Acids Res. **8**, 3467-3480.

LOENEN, W.A.M., and BLATTNER, F.R. (1983). Lambda Charon vectors (Ch32, 33, 34 and 35) adapted for DNA cloning in recombination-deficient hosts. Gene **26**, 171-179.

LU, H.S., YUAN, P.M., and GRACY, R.W. (1984). Primary structure of human triosephosphate isomerase. J. Biol. Chem. **259**, 11958-11968.

MANIATIS, T., FRITSCH, E.F., and SAMBROOK, J. (1982). Molecular cloning. A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

MANLEY, J.L., FIRE, A., CANO, A., SHARP, P.A., and GEFTER, M.L. (1980). DNA-dependent transcription of adenovirus genes in a soluble whole-cell extract. Proc. Natl. Acad. Sci. USA **77**, 3855-3859.

MAQUAT, L.E., CHILCOTE, R., and RYAN, P.M. (1985). Human triosephosphate isomerase cDNA and protein structure: studies of triosephosphate isomerase deficiency in man. J. Biol. Chem. **260**, 3748-3753.

MARCHIONNI, M., and GILBERT, W. (1986). The triosephosphate isomerase gene from maize: introns antedate the plant-animal divergence. Cell **46**, 133-141.

MARTINI, G., TONIOLO, D., VULLIAMY, T., LUZZATTO, L., DONO, R., VIGLIETTO, G., PAONESSA, G., D'URSO, M., and PERSICO, M.G. (1986). Structural analysis of the X-linked gene encoding human glucose 6-phosphate dehydrogenase. EMBO J. **5**, 1849-1855.

McBRIDE, O.W., PIRTLE, I.L., and PIRTLE, R.M. (1989). Localization of three DNA segments encompassing tRNA genes to human chromosomes 1,5, and 16: Proposed mechanism and significance of tRNA gene dispersion. Genomics (in press).

McKNIGHT, G.L., O'HARA, P.J., and PARKER, M.L. (1986). Nucleotide sequence of the triosephosphate isomerase gene from *Aspergillus nidulans*: implications for a differential loss of introns. Cell **46**, 143-147.

MITCHELL, P.J., and TJIAN, R. (1989). Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. Science **245**, 371-378.

MIYAKE, T., MIGITA, K., and SAKAKI, Y. (1985). Some *KpnI* family members are associated with the *Alu* family in the human genome. Nucl. Acids Res. **11**, 6837-6846.

MIYAMOTO, M.M., SLIGHTOM, J.L., and GOODMAN, M. (1987). Phylogenetic relations of humans and African apes from DNA sequences in the ψη-globin region. Science **238**, 369-373.

MIYATA, T., and YASUNAGA, T. (1981). Rapidly evolving mouse α-globin-related pseudo gene and its evolutionary history. Proc. Natl. Acad. Sci. USA **78**, 450-453.

MURPHY, S., DILIEGRO, C., and MELLI, M. (1987). The *in vitro* transcription of the 7SK RNA gene by RNA polymerase III is dependent only on the presence of an upstream promoter. Cell **51**, 81-87.

NAIDU, J.M., TURNER, T.R., and MOHRENWEISER, H.W. (1984). Thermostability characteristics of glucosephosphate and triosephosphate isomerase in erythrocytes from several species. Comp. Biochem. Physiol. **79B**, 211-217.

OCHMAN, H., and WILSON, A.C. (1987). Evolution in bacteria: evidence for a universal substitution rate in cellular genomes. J. Mol. Evol. **26**, 74-86.

OLD, S.E. (1988). Primate triosephosphate isomerase isozymes and gene structure. Ph.D. Thesis. University of Michigan.

OLD, S.E., LANDA, L.E., and MOHRENWEISER, H.W. (1989).

Hominoid triosephosphate isomerase: regulation of expression of the proliferation specific isozyme. Mol. Cell. Biochem. 89, 73-85.

OLD, S.E., and MOHRENWEISER, H.W. (1988). Nucleotide sequence of the triosephosphate isomerase gene from *Macaca mulatta*. Nucl. Acids Res. **16**, 9055.

ONO, M. (1986). Molecular cloning and long terminal repeat sequences of human endogenous retrovirus genes related to types A and B retrovirus genes. J. Virol. **58**, 937-944.

ONO, M., YASUNAGA, T., MIYATA, T., and USHIKUBO, H. (1986). Nucleotide sequence of human endogenous retrovirus genome related to the mouse mammary tumor virus genome. J. Virol. **60**, 589-598.

PEARSON, W.R., and LIPMAN, D.J. (1988). Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. USA **77**, 2444-2448.

PERLER, F., EFSTRATIADIS, A., LOMEDICO, P., GILBERT, W., KOLODNER, R., and DODGSON, J. (1980). The evolution of genes: the chicken preproinsulin gene. Cell **20**, 555-565.

PETERSON, R.C. (1987). Sequence and transcription of $tRNA^{Val}$ gene from *Xenopus laevis*. Biochem. Biophys. Acta **908**, 81-89.

PICHERSKY, E., GOTTLIEB, L.D., and HESS, J.F. (1984). Nucleotide sequence of the triose phosphate isomerase gene of *Escherichia coli*. Mol. Gen. Genet. **195**, 314-320.

PIPER, P.W. (1975). The primary structure of the major cytoplasmic valine tRNA of mouse myeloma cells. Eur. J. Biochem. **51**, 295-304.

PIRTLE, R.M., PIRTLE, I.L., AND INOUYE, M. (1980). Messenger ribonucleic acid of the lipoprotein of the *Escherichia coli* outer membrane: I. Nucleotide sequence of

the 3'-terminus and sequences of oligonucleotides derived from complete digests of the mRNA. J. Biol. Chem. **255**, 199-209.

PIRTLE, I.L., SHORTRIDGE, R.D., and PIRTLE, R.M. (1986). Nucleotide sequence and transcription of a human glycine $tRNA_{GCC}$ gene and nearby pseudogene. Gene **43**, 155-167.

QUENTIN, Y. (1988). The *Alu* family developed through successive waves of fixation closely connected with primate lineage history. J. Mol. Evol. **27**, 194-202.

RABA, M., LIMBURG, K., BURGHAGEN, M., KATZE, J.R., SIMSEK, M., HECKMAN, J.E., RAJBHANDARY, U.L., and GROSS, H.J. (1979). Nucleotide sequence of three isoaccepting lysine tRNAs from rabbit liver and SV40-transformed mouse fibroblasts. Eur. J. Biochem. **97**, 305-318.

RACKWITZ, H., ZEHETNER, G., FRISCHAUF, A., and LEHRACH, H. (1984). Rapid restriction mapping of DNA cloned in lambda phage vectors. Gene **30**, 195-200.

REED, K.C., and MANN, D.A. (1985). Rapid transfer of DNA from agarose gels to nylon membranes. Nuc. Acids Res. **13**, 7207-7221.

RICH, A., NORDHEIM, A., and WANG, A.H.-J. (1984). Thechemistry and biology of left-handed Z-DNA. Annu. Rev. Biochem. **53**, 791-846.

RICH, A., and RAJBHANDARY, U.L. (1976). Transfer RNA:molecular structure, sequence and properties. Annu. Rev. Biochem. **45**, 805-860.

RIGBY, P.W.J., DIECKMANN, M., RHODES, C., and BERG, P. (1977). Labeling deoxyribonucleic acid to high specific activity in vitro by nick translation with DNA polymerase I. J. Mol. Biol. **113**, 237-251.

ROGERS, P.A., BRENTON, D.P., and HOPKINSON, D.A. (1980).

Changes in the activity and isozyme patterns of glycolytic enzymes during stimulation of normal human lymphocytes with phytohaemagglutinin. Ann. Hum. Genet. **43**, 213-226.

ROSEN, A., and DANIEL, V. (1988). Nucleotide sequence and transcription of a rat $tRNA^{Phe}$ gene and a neighboring *Alu*-like element. Gene **69**, 275-285.

ROY, K.L., COOKE, H., and BUCKLAND, R. (1982). Nucleotide sequence of a segment of human DNA containing three tRNA genes. Nucl. Acids Res. **10**, 7313-7322.

RUBIN, C.M., HOUCK, C.M., DEININGER, P.L., FRIEDMANN, T., and SCHMID, C.W. (1980). Partial nucleotide sequence of the 300-nucleotide interspersed repeated human DNA sequences. Nature **284**, 372-374.

RUBINSON, H., VODOVAR, M., MEIENHOFER, M.C., and DREYFUS, J.C. (1971). A unique electrophoretic pattern of triosephosphate isomerase in human cultured fibroblast. FEBS Lett. **13**, 290-292.

RUSSELL, P.R. (1985). Transcription of the triose-phosphate-isomerase gene of *Schizosaccharomyces pombe* initiates from a start point different from that in *Saccharomyces cerevisiae*. Gene **40**, 125-130.

RUSSO, T., OLIVA, A., DUILIO, A., AMMENDOLA, R., COSTANZO, F., ZANNINI, M., and CIMINO, F. (1987). The transcriptional efficiency of clustered tRNA genes is affected by their position within the cluster. Biochem. Biophys. Res. Commun. **149**, 1118-1124.

SANGER, F., COULSON, A.R., BARRELL, B.G., SMITH, A.J.H., and ROE, B.A. (1980). Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. J. Mol. Biol. **143**, 161-178.

SANGER, F., NICKLEN, S., and COULSON, A.R. (1977). DNA sequencing with chain terminating inhibitors. Proc. Natl.

Acad. Sci. USA **74**, 5463-5467.

SANTOS, T., and ZASLOFF, M. (1981). Comparative analysis of human chromosomal segments bearing nonallelic dispersed $tRNA^{Met}$ genes. Cell **23**, 699-709.

SAWADA, I., WILLARD, C., SHEN, C.-K.J., CHAPMAN, B., WILSON, A.C., and SCHMID, C.W. (1985). Evolution of *Alu* family repeats since the divergence of human and chimpanzee. J. Mol. Evol. **22**, 316-322.

SCARPULLA, R.C. (1985). Association of a truncated cytochrome C processed pseudogene with a similarly truncated member from a long interspersed repeat family of rat. Nucl. Acids Res. **13**, 763-775.

SCHEIDERIT, C., and BEATO, M. (1984). Contacts between hormone receptor and DNA double helix within a glucocorticoid regulatory element of mouse mammary tumor virus. Proc. Natl. Acad. Sci. USA **81**, 3029-3033.

SCOTT, A.F., SCHMECKPEPER, B.J., ABDELRAZIK, M., COMEY, C.T., O'HARA, B., ROSSITER, J.P., COOLEY, T., HEATH, P., SMITH, K.D., and MARGOLET, L. (1987). Origin of the human L1 elements: proposed progenitor genes deduced from a consensus DNA sequence. Genomics **1**, 113-125.

SEKIYA, T., NISHIZAWA, R., MATSUDA, K., TAYA, Y., and NISHIMURA, S. (1982). A rat tRNA gene cluster containing the genes for $tRNA^{Pro}$ and $tRNA^{Lys}$. Analysis of nucleotide sequences of the genes and the surrounding regions. Nuc. Acids Res. **10**, 6411-6419.

SERFLING, E., JASIN, M., and SCHAFFNER, W. (1985). Enhancers and eukaryotic gene-transcription. Trends Genet. **1**, 224-230.

SHAPIRO, L.H., VENTA, P.J., and TASHIAN, R.E. (1987). Molecular analysis of G+C-rich upstream sequences

regulating transcription of the human carbonic anhydrase II gene. Mol. Cell. Biol. **7**, 4589-4593.

SHELLEY, C.S., SHARPE, C.R., BARALLE, F.E., and SHOULDERS, C.C. (1985). Comparison of the human apolipoprotein genes. Apo AII presents a unique functional intron-exon junction. J. Mol. Biol. **186**, 43-51.

SHORTRIDGE, R.D., JOHNSON, G.D., CRAIG, L.C., PIRTLE, I. L., and PIRTLE, R.M. (1989). A human tRNA gene heterocluster encoding threonine, proline and valine tRNAs. Gene **79**, 309-324.

SHORTRIDGE, R.D., PIRTLE, I.L., and PIRTLE, R.M. (1986). IBM microcomputer programs that analyze DNA sequences for tRNA genes. Computer Applications in the Biosciences **2**, 13-17.

SILVERMAN, S., GILLAM, I.C., TENER, G.M., and SÖLL, D. (1979). The nucleotide sequence of lysine $tRNA_2$ from *Drosophila*. Nucl. Acids Res. **6**, 435-442.

SLIGHTOM, J.L., BLECHL, A.E., and SMITHIES, O. (1980). Human fetal $G\gamma$- and $A\gamma$-globin genes: Complete nucleotide sequences suggest that DNA can be exchanged between these duplicated genes. Cell **21**, 627-638.

SOLLNER-WEBB, B. (1988). Surprises in polymerase III transcription. Cell **52**, 153-154.

SOUTHERN, E.M. (1979). Gel electrophoresis of restriction fragments. Meth. Enzymol. **68**, 152-176.

STRAUS, D., and GILBERT, W. (1985). Genetic engineering in the precambrian: structure of the chicken triosephosphate isomerase gene. Mol. Cell. Biol. **5**, 3497-3506.

SWINKLES, B.W., GIBSON, W.C., OSINGA, K.A., KRAMER, R., VEENEMAN, G.H., VAN BOOM, J.H., and BORST, P. (1986).

Characterization of the gene for the microbody (glycosomal) triosephosphate isomerase of *Trypansoma brucei*. EMBO J. **5**, 1291-1298.

TEMIN, H.M. (1981). Structure, variation and synthesis of retrovirus long terminal repeat. Cell **27**, 1-3.

TYLER, B.M. (1987). Transcription of *Neurospora crassa* 5S rRNA genes requires a TATA box and three internal elements. J. Mol. Biol. **196**, 801-811.

VANIN, E.F. (1985). Processed pseudogenes: characteristics and evolution. Annu. Rev. Genet. **19**, 253-272.

VARMUS, H.E. (1982). Form and function of retroviral proviruses. Science **216**, 812-820.

WANG, L. (1989). Physical mapping of human transfer RNA gene clusters. M.S. Thesis. University of North Texas.

WEIHER, H., KONIG, M., and Gruss, P. (1983). Multiple point mutations affecting the simian virus 40 enhancer. Science **219**, 626-631.

WEIL, P.A., SEGALL, J., HARRIS, B., NG, S.Y., and ROEDER, R.G. (1979). Faithful transcription of eukaryotic genes by RNA polymerase III in systems reconstituted with purified DNA templates. J. Biol. Chem. **254**, 6163-6173.

WEINER, A.M., DEININGER, P.L., and EFSTRATIADIS, A. (1986). Nonviral retroposons: genes, pseudogenes, and transposable elements genterated by the reverse flow of genetic information. Annu. Rev. Biochem. **55**, 631-661.

WELLS, R.D. (1988). Unusual DNA structures. J. Biol. Chem. **263**, 1095-1098.

WHITE, M.B., SHEN, A.L., WORD, C.J., TUCKER, P.W., and BLATTNER, F.R. (1985). Human immunoglobulin D: genomic sequence of the delta heavy chain. Science **228**, 733-737.

WILLARD, C., WONG, E., HESS, J.F., SHEN, C.-K.J., CHAPMAN, B., WILSON, A.C., and SCHMID, C.W. (1985). Comparison of human and chimpanzee ζ1 globin genes. J. Mol. Evol. **22**, 309-315.

YAMAMOTO, K.R. (1985). Steroid receptor regulated transcription of specific genes and gene networks. Annu. Rev. Genet. **19**, 209-252.

YANISCH-PERRON, C., VIEIRA, J., and MESSING, J. (1985). Improved M13 phage cloning vectors and host strains: Nucleotide sequences of the M13mp18 and pUC vectors. Gene **33**, 103-119.

YOUNG, L.S., TAKAHASHI, N., and SPRAGUE, K.U. (1986). Upstream sequences confer distinctive transcriptional properties on genes encoding silkgland-specific $tRNA^{Ala}$. Proc. Natl. Acad. Sci. USA **83**, 374-378.

YUAN, P.M., DEWAN, R.N., ZAUN, M., THOMPSON, R.E., and GRACY, R.W. (1979). Isolation and characterization of triosephosphate isomerase isozymes from human placenta. Arch. Biochem. Biophys. **198**, 42-52.

ZHANG, H., SCHOLL, R., BROWSE, J., and SOMERVILLE, C. (1988). Double-stranded DNA sequencing as a choice for DNA sequencing. Nucl. Acids Res. **16**, 1220.