Isolation and characterization of the human tyrosine aminotransferase gene

Ruth Rettenmeier, Ernst Natt+, Hanswalter Zentgraf[1] and Gerd Scherer*
Institute of Human Genetics, Alberstraße 11, D7800 Freiburg i.Br. and [1]Institute of Virology, German Cancer Research Center, Im Neuenheimer Feld 280, D6900 Heidelberg, FRG

Received March 28, 1990; Revised and Accepted June 1, 1990 EMBL accession nos X52507–X52520 (incl.)

ABSTRACT

Structure and sequence of the human gene for tyrosine aminotransferase (TAT) was determined by analysis of cDNA and genomic clones. The gene extends over 10.9 kbı and consists of 12 exons giving rise to a 2,754 nucleotide long mRNA (excluding the poly(A)tail). The human TAT gene is predicted to code for a 454 amino acid protein of molecular weight 50,399 dalton. The overall sequence identity within the coding region of the human and the previously characterized rat TAT genes is 87% at the nucleotide and 92% at the protein level. A minor human TAT mRNA results from the use of an alternative polyadenylation signal in the 3′ exon which is present but not used at the corresponding position in the rat TAT gene. The non-coding region of the 3′ exon contains a complete Alu element which is absent in the rat TAT gene but present in apes and old world monkeys. Two functional glucocorticoid response elements (GREs) reside 2.5 kb upstream of the rat TAT gene. The DNA sequence of the corresponding region of the human TAT gene shows the distal GRE mutated and the proximal GRE replaced by Alu elements.

INTRODUCTION

Tyrosine aminotransferase (TAT, EC 2.6.1.5) is a liver-specific enzyme that converts tyrosine to p-hydroxyphenylpyruvate in a pyridoxal phosphate-dependent transamination reaction. Regulation of TAT enzyme activity by glucocorticoids, insulin and glucagon has been studied extensively in the rat (for review see 1). Induction of the enzyme by glucocorticoid hormones and cAMP is a consequence of direct transcriptional activation of the gene (2, 3). Two glucocorticoid response elements (GREs) activating transcription in a cooperative manner have been identified 2.5 kb upstream of the transcription initiation site of the rat TAT gene (4), while an additional regulatory region essential for full glucocorticoid induction resides at −5.4 kb (5). TAT genes have been cloned from both rat (6) and mouse (7), and the complete cDNA sequence for rat TAT is known (8). Human TAT enzyme has been purified and shown to be a dimer of molecular weight 95,500 dalton (9). Induction of human TAT by glucocorticoids and cAMP has been demonstrated in fetal liver organ cultures (10, 11).

Deficiency in TAT activity causes tyrosinemia II (Richner-Hanhart syndrome), a disease showing autosomal recessive inheritance (for review see 12). From an interest in the molecular genetics of this disorder, and in the hormonal regulation of the human TAT gene, we have isolated genomic and cDNA clones for human TAT. These cloned sequences have been used by us to assign the TAT locus to the long arm of human chromosome 16 (13), to identify restriction fragment length polymorphisms for use in linkage analysis and diagnosis in tyrosinemia II (14, 15), and to characterize deletions and point mutations within the TAT gene in tyrosinemia II patients (16; E. Natt & G. Scherer, unpublished). In this report, we present the complete exon/intron structure and exon sequence of the human TAT gene. High sequence similarity is found when exonic and 5′ -flanking regions of the human and rat TAT genes are compared, but the human sequence corresponding to the −2.5 kb rat GRE region has highly diverged. Features unique to the human TAT gene are the production of two mRNA species and the presence of an Alu element in the 3′ exon.

MATERIAL AND METHODS

Screening of human genomic and cDNA libraries

Two human genomic libraries constructed on Sau3A partial digests were used, one prepared from placenta DNA (a gift of N. Blin, Homburg) in the phage vector EMBL3 according to standard protocols (17), the other made from DNA of a male Burkitt lymphoma in the vector EMBL3A (library kindly provided by M. Lipp, Munich). About 8×10^5 recombinants from both libraries were screened (18) with pcTAT-3 (19), a rat TAT cDNA clone covering exons F to K of the rat TAT gene (6). After three rounds of plaque purification, one clone was isolated from each library, λhTAT1 from the EMBL3 library and λhTAT2 from the EMBL3A library.

The human adult liver cDNA library used (kindly provided by T. Chandra and S. Woo, Houston; 20) consists of cDNA

* To whom correspondence should be addressed

+ Present address: Plant Biology Laboratory, The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, CA 92037, USA

fragments cloned with Eco RI linkers into the EcoRI cloning site of phage vector λgt11. About 6×10^5 recombinants from this library were screened sequentially with the following probes: with pcTAT-3 mentioned above, with pUTAT-EH 0.94, a genomic clone spanning exons B and C of the rat TAT gene (6), and with an 0.25 kb EcoRI/PstI fragment from the 5′ end of the human cDNA clone phcTAT2-16.

DNA sequence analysis

Overlapping restriction fragments of the cDNA clones phcTAT2-16 and phcTAT3a-6 were subcloned into the sequencing vectors M13mp18/mp19 and sequenced from both strands by the dideoxynucleotide chain termination method (21) using [α-^{35}S]dATP (Amersham) and a modified T7 DNA polymerase (Sequenase, United States Biochemical Corp.). Exon/intron junctions and the 5′ and 3′ flanking region of the TAT gene were sequenced accordingly from subclones of the genomic clone λhTAT1. In addition, some of the DNA fragments were also sequenced by the chemical method of Maxam and Gilbert (22). Sequence data were handled by the IBI/Pustell DNA And Protein Sequence Analysis System purchased from IRL Press Ltd. (Oxford, England).

Reverse transcriptase–PCR

Total RNA from human liver (3 μg) was reverse transcribed in 20 μl using random hexamer primers (Pharmacia) essentially as described (23). Of this, 5 μl was added to a 60 μl PCR reaction containing 10 mM Tris-HCl, pH 8.4, 50 mM KCl, 3 mM $MgCl_2$, 0.2 mM each dNTP, 0.05% Nonidet P-40, 0.05% Tween 20, 25 μg/ml BSA, 10 pmol each primer, and 1 unit Taq polymerase (Amersham). DNA was amplified by 40 cycles of 94°C for 1 min, 58°C for 1 min, and 72°C for 2 min in a Techne PHC-1. PCR primers used correspond to nucleotides +36 to +56 in exon A (forward primer), and to the complement of nucleotides +304 to +326 in exon B (reverse primer). The 291 bp PCR product was cloned into SmaI linearized M13mp18 and sequenced from both orientations.

Northern blot analysis and primer extension analysis

Human liver samples (a gift of G. Kovacs, Hannover) and rat liver, frozen in liquid nitrogen, were ground in a mortar, and total RNA was isolated from the resulting powder by the lithium chloride/urea procedure (24). Poly (A)$^+$RNA was isolated using Hybond-mAP paper (Amersham). RNAs were transferred to Gene Screen nylon membranes (NEN) after electrophoretic separation in 1.2% agarose-formaldehyde gels. Membranes were hybridized with probes labelled by the random primer technique (25). For quantitation of bands, autoradiographs were exposed to be in linear range and scanned with a Shimadzu dual wave length scanner (model CS-930).

For primer extension, a 5′ end-labeled oligonucleotide (0.5 pmol, 1×10^6 cpm ^{32}P) complementary to nucleotides +49 to

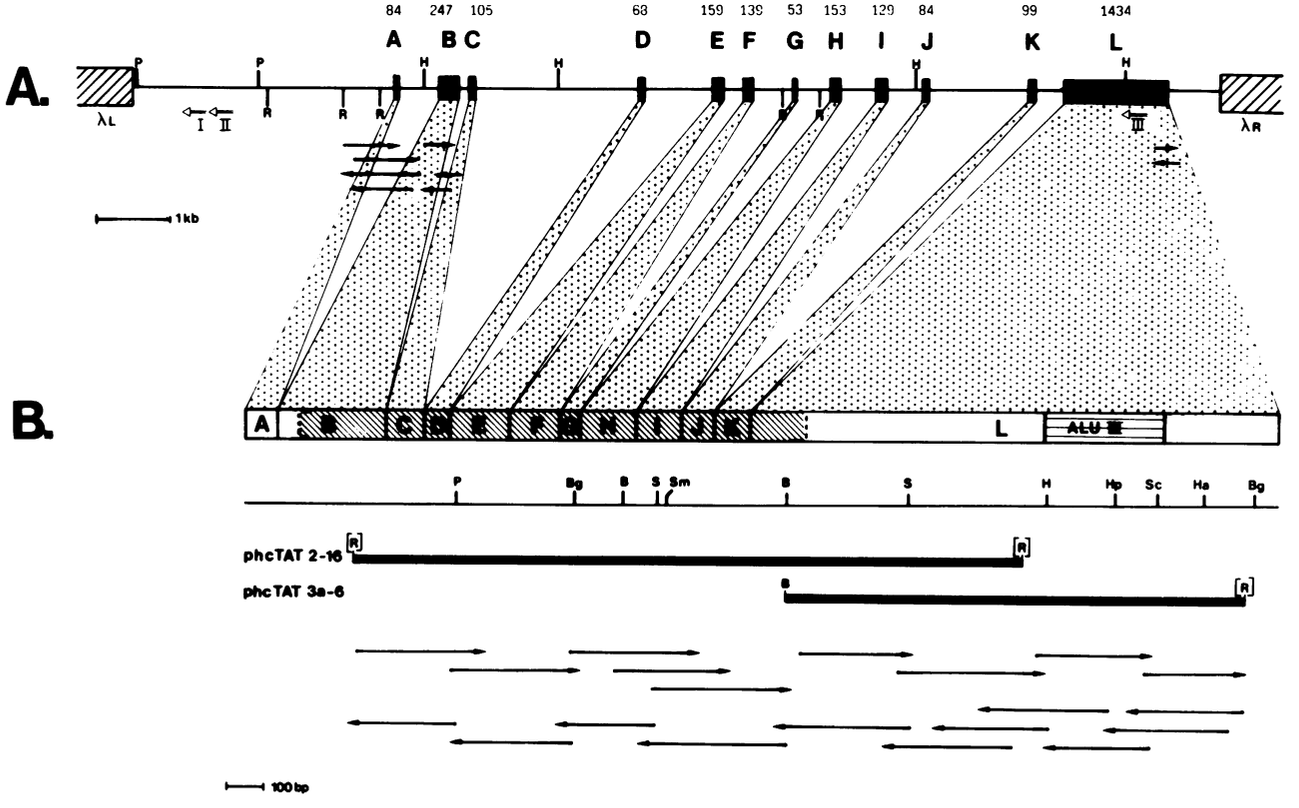
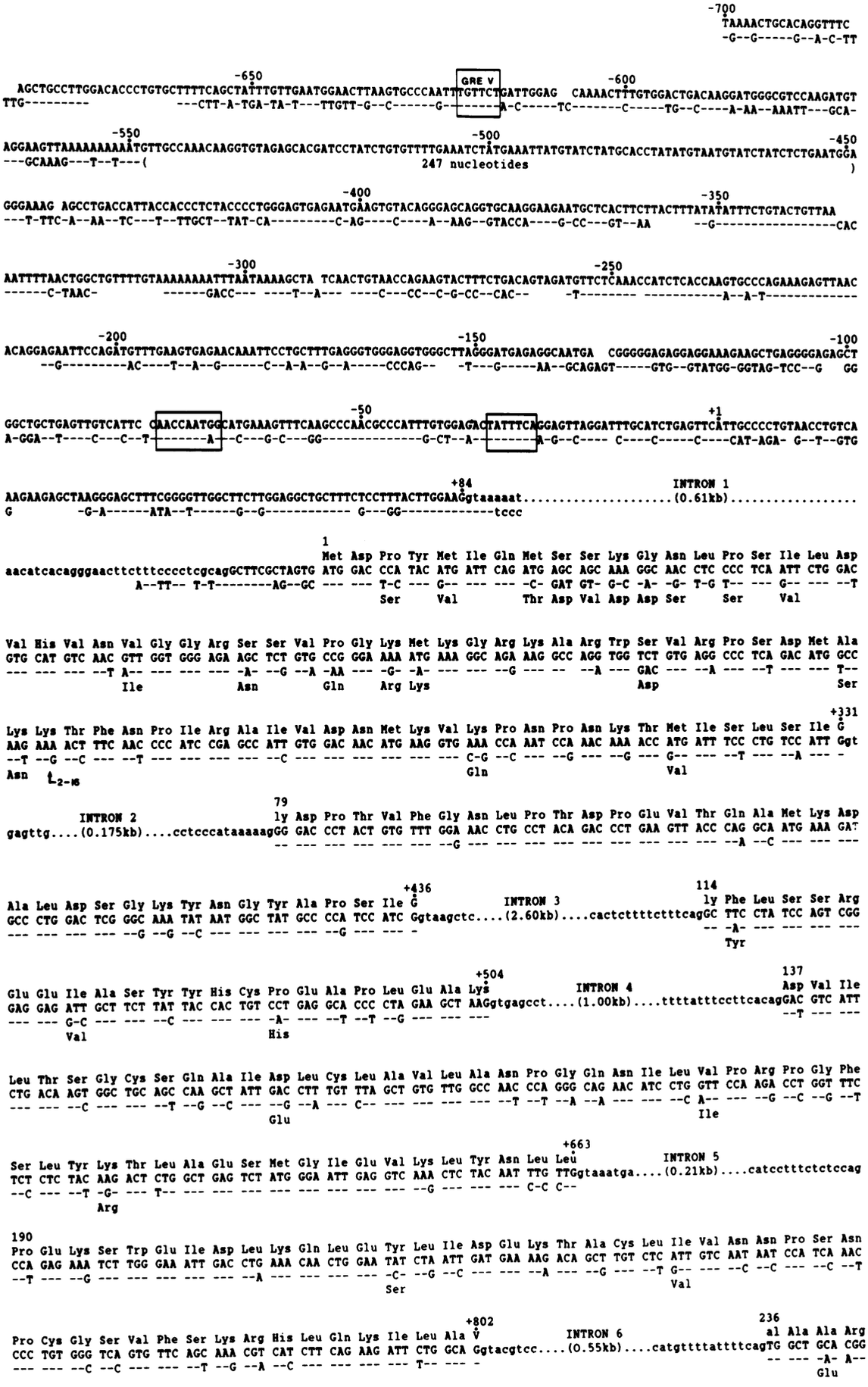


Figure 1: Organization of the human TAT gene and sequencing strategy of TAT cDNA clones. A:Exon/intron structure and restriction map of clone λhTAT1. Size in nucleotides of exons A to L (black bars) as determined by DNA sequence analysis (Fig. 2). Intron sizes are given in Fig. 2. I, II, III, Alu elements, their orientation indicated by arrows. λL and λR denote left and right arms of the EMBL3 vector, respectively. Arrows indicate direction and extent of genomic DNA sequences determined. B:Exon map and restriction endonuclease map of human TAT cDNA. The area representing the coding sequence is hatched, the Alu element in exon L is indicated. Sequences covered by cDNA clones phcTAT2-16 and phcTAT3a-6 are indicated by black bars. Direction and extent of cDNA sequences determined are shown by arrows. Restriction sites in A and B are designated as follows: B, BamHI; Bg, BglII; H, HindIII; Ha, HaeIII; Hp, HpaII; P, PstI (in A, only the PstI sites flanking the 1.6 kb PstI fragment from Fig. 4 are indicated); R, EcoRI; [R], EcoRI linker; S, SstI; Sm, SmaI; Sc, ScaI.


-700
TAAAACTGCACAGGTTTC
-650
GRE V
-600
AGCTGCCTTGGACACCCTGTGCTTTTCAGCTATTTGTTGAATGGGAACTTAAGTGCCCAATTTGTTCTGATTGGAG CAAAACTTTGTGGACTGACAAGGATGGGCGTCCAAGATGT
-550
-500
-450
AGGAAGTTAAAAAAAAAATGTTGCCAAACAAGGTGTAGAGCACGATCCTATCTGTGTTTTGAAATCTATGAAATTATGTATCTATGCACCTATATGTAATGTATCTATCTCTGAATGGA
247 nucleotides
-400
-350
GGGAAAG AGCCTGACCATTACCACCCTCTACCCCTGGGAGTGAGAATGAAGTGTACAGGGAGCAGGTGCAAGGAAGAATGCTCACTTCTTACTTTATATATTTCTGTACTGTTAA
-300
-250
AATTTTAACTGGCTGTTTTGTAAAAAAAATTTAATAAAAGCTA TCAACTGTAACCAGAAGTACTTTCTGACAGTAGATGTTCTCAAACCATCTCACCAAGTGCCCAGAAAGAGTTAAC
-200
-150
-100
ACAGGAGAATTCCAGATGTTTGAAGTGAGAACAAATTCCTGCTTTGAGGGTGGGAGGTGGGCTTAGGGATGAGAGGCAATGA CGGGGGAGAGGAGGAAAGAAGCTGAGGGGAGAGCT
-50
+1
GGCTGCTGAGTTGTCATTC CAACCAATGGCATGAAAGTTTCAAGCCCAACGCCCATTTGTGGAGACTATTTCAGGAGTTAGGATTTGCATCTGAGTTCATTGCCCCTGTAACCTGTCA
+84
INTRON 1
AAGAAGAGCTAAGGGAGCTTTCGGGGTTGGCTTCTTGGAGGCTGCTTTCTCCTTTACTTGGAAGgtaaaaat.....(0.61kb).....
1
Met Asp Pro Tyr Met Ile Gln Met Ser Ser Lys Gly Asn Leu Pro Ser Ile Leu Asp
aacatcacagggaacttctttcccctcgcagGCTTCGCTAGTG ATG GAC CCA TAC ATG ATT CAG ATG AGC AGC AAA GGC AAC CTC CCC TCA ATT CTG GAC
Ser Val Thr Asp Val Asp Asp Ser Ser Val
Val His Val Asn Val Gly Gly Arg Ser Ser Val Pro Gly Lys Met Lys Gly Arg Lys Ala Arg Trp Ser Val Arg Pro Ser Asp Met Ala
GTG CAT GTC AAC GTT GGT GGG AGA AGC TCT GTG CCG GGA AAA ATG AAA GGC AGA AAG GCC AGG TGG TCT GTG AGG CCC TCA GAC ATG GCC
Ile Asn Gln Arg Lys Asp Ser
+331
Lys Lys Thr Phe Asn Pro Ile Arg Ala Ile Val Asp Asn Met Lys Val Lys Pro Asn Pro Asn Lys Thr Met Ile Ser Leu Ser Ile G
AAG AAA ACT TTC AAC CCC ATC CGA GCC ATT GTG GAC AAC ATG AAG GTG AAA CCA AAT CCA AAC AAA ACC ATG ATT TCC CTG TCC ATT Ggt
Asn L2-16
Gln Val
79
INTRON 2
ly Asp Pro Thr Val Phe Gly Asn Leu Pro Thr Asp Pro Glu Val Thr Gln Ala Met Lys Asp
gagttg....(0.175kb)....cctcccataaaaagGG GAC CCT ACT GTG TTT GGA AAC CTG CCT ACA GAC CCT GAA GTT ACC CAG GCA ATG AAA GAT
+436
114
Ala Leu Asp Ser Gly Lys Tyr Asn Gly Tyr Ala Pro Ser Ile G
INTRON 3
ly Phe Leu Ser Ser Arg
GCC CTG GAC TCG GGC AAA TAT AAT GGC TAT GCC CCA TCC ATC Ggtaagctc....(2.60kb)....cactcttttctttcagGC TTC CTA TCC AGT CGG
Tyr
+504
137
Glu Glu Ile Ala Ser Tyr Tyr His Cys Pro Glu Ala Pro Leu Glu Ala Lys
INTRON 4
Asp Val Ile
GAG GAG ATT GCT TCT TAT TAC CAC TGT CCT GAG GCA CCC CTA GAA GCT AAGgtgagcct....(1.00kb)....ttttatttccttcacagGAC GTC ATT
Val His
Leu Thr Ser Gly Cys Ser Gln Ala Ile Asp Leu Cys Leu Ala Val Leu Ala Asn Pro Gly Gln Asn Ile Leu Val Pro Arg Pro Gly Phe
CTG ACA AGT GGC TGC AGC CAA GCT ATT GAC CTT TGT TTA GCT GTG TTG GCC AAC CCA GGG CAG AAC ATC CTG GTT CCA AGA CCT GGT TTC
Glu Ile
+663
Ser Leu Tyr Lys Thr Leu Ala Glu Ser Met Gly Ile Glu Val Lys Leu Tyr Asn Leu Leu
INTRON 5
TCT CTC TAC AAG ACT CTG GCT GAG TCT ATG GGA ATT GAG GTC AAA CTC TAC AAT TTG TTGgtaaatga....(0.21kb)....catcctttctctccag
Arg
190
Pro Glu Lys Ser Trp Glu Ile Asp Leu Lys Gln Leu Glu Tyr Leu Ile Asp Glu Lys Thr Ala Cys Leu Ile Val Asn Asn Pro Ser Asn
CCA GAG AAA TCT TGG GAA ATT GAC CTG AAA CAA CTG GAA TAT CTA ATT GAT GAA AAG ACA GCT TGT CTC ATT GTC AAT AAT CCA TCA AAC
Ser Val
+802
236
Pro Cys Gly Ser Val Phe Ser Lys Arg His Leu Gln Lys Ile Leu Ala V
INTRON 6
al Ala Ala Arg
CCC TGT GGG TCA GTG TTC AGC AAA CGT CAT CTT CAG AAG ATT CTG GCA Ggtacgtcc....(0.55kb)....catgttttattttcagTG GCT GCA CGG
Glu

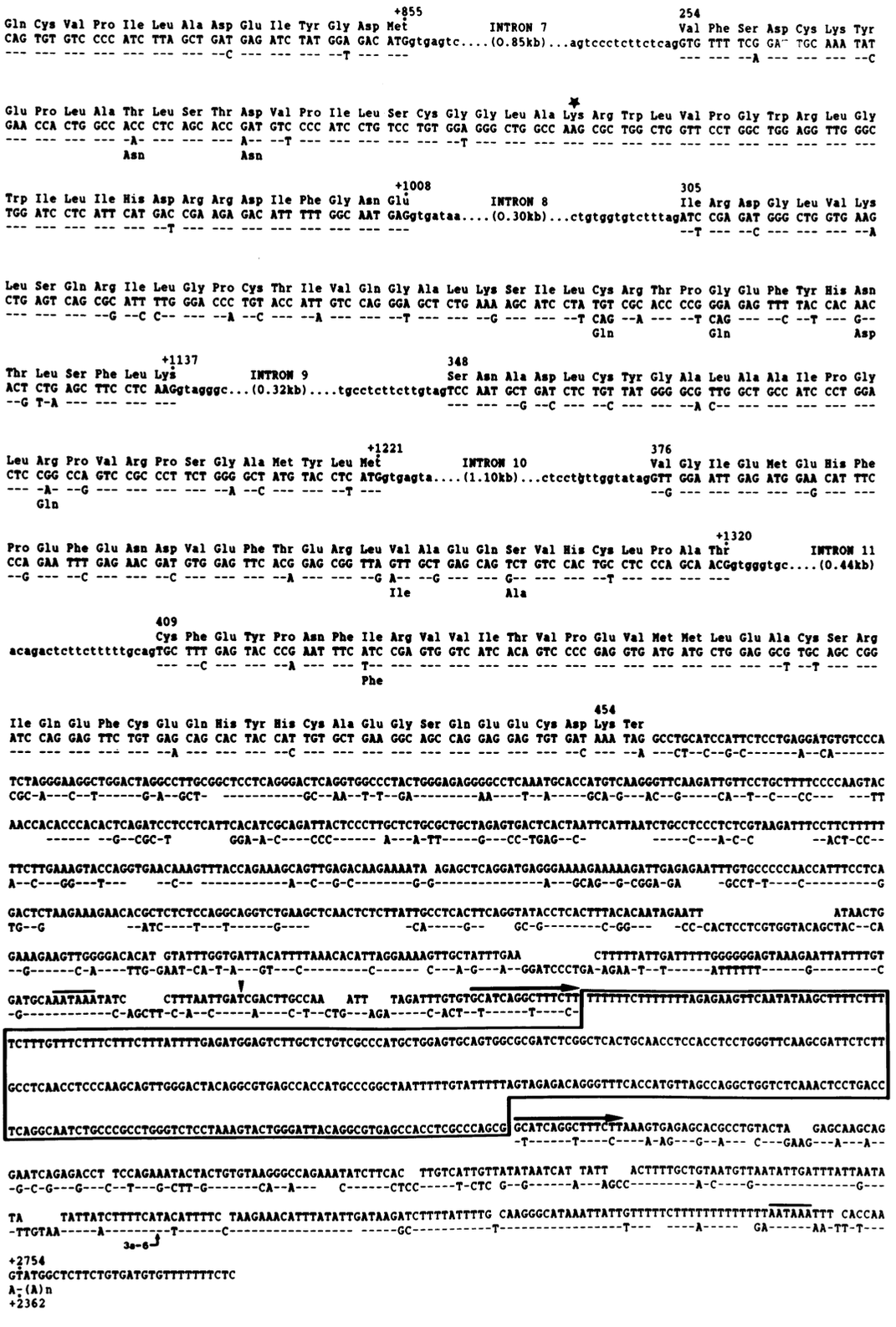


Figure 2: Sequence of the 5′ flanking region and of exons A to L of the human TAT gene and comparison with the rat TAT sequence. Human TAT sequences are given in the upper lane, rat TAT sequences (6, 8, 28) in the lower lane. Sequences are aligned to maximize for identities, indicated by dashes in the rat sequence, while empty spaces denote gaps. Nucleotide position +1 corresponds to the transcription start site and position +2754 to the 3′ end of the rat TAT cDNA at position +2362 (8). Nucleotides 5′ to position +1 are given negative numbers. Sequences at the end of introns are in lower case, and the intron lengths determined by heteroduplex analysis and restriction mapping are given. The 5′ end in exon B of phcTAT2-16 and the 3′ end in exon L of phcTAT3a-6 are indicated. The predicted amino acid sequence for human TAT is shown above the nucleotide sequences, while amino acid differences in rat TAT are written below. The lysine residue at codon 280 that binds the cofactor pyridoxal phosphate in rat TAT (30) is marked by an asterisk. Ter, stop codon. Numbering at end of exons is for nucleotides and at beginning of exons C to L for amino acids. TATA- and CAAT-like sequences are boxed, as is the TGTTCT motif of the GRE V element. Two AATAAA signal sequences in the 3′ non-coding region are overlined, and the polyadenylation site of cDNA clone phcTAT2-33 is indicated by an arrowhead (see Fig. 3). The Alu element in exon L is boxed, the flanking direct repeats are marked by arrows.

+65 in exon A was hybridized to 50 μg human liver total RNA at 50 °C for 2 h, followed by incubation with reverse transcriptase at 42 °C for 1 h, exactly as described (26). A 600 bp genomic EcoRI/HindIII fragment spanning exon A was cloned in M13mp18 and served as template for dideoxynucleotide sequencing (21) using unlabeled dNTPs and the same 5′ end-labeled oligonucleotide primer. The sequence reaction products were run in parallel with the primer-extended fragments on a 12% polyacrylamide, 7 M urea sequencing gel.

RESULTS AND DISCUSSION

Structure and sequence of the human TAT gene and comparison with the rat TAT gene sequence

Using the partial rat cDNA clone pcTAT-3 (19) as probe, we isolated two phage clones from two different human genomic libraries. The restriction map of one of the clones, λhTAT1, which contains the entire human TAT gene, is shown in Fig. 1A. This clone carries a 15.2 kb insert with 3.5 kb of 5′ flanking DNA, 10.9 kb of gene sequences, and 0.8 kb of 3′ flanking DNA. The second clone, λhTAT2, carries an insert of 15.3 kb and extends a further 2.5 kb in the 3′ direction (not shown). From a human adult liver cDNA library, a total of 15 cDNA clones were isolated and characterized in detail. Analysis of heteroduplex molecules formed between cDNA and genomic clones (not shown; 27) and comparison with the exon/intron maps of rat and mouse TAT genes (6, 7) revealed that cDNA clones phcTAT2-16 and phcTAT 3a-6 represent most of the human TAT cDNA. To obtain the complete sequence of the human TAT gene exons, appropriate fragments from these cDNA clones were sequenced according to the sequencing strategy shown in Fig. 1 B. In addition, genomic sequences from clone λhTAT1 were obtained (Fig. 1 A) from 700 bp of 5′ flanking sequence, from 3′ flanking sequences, and from restriction fragments spanning the exon/intron boundaries.

The resulting human TAT sequences are presented in Fig. 2 and compared with sequences from the promoter region of the rat TAT gene (6, 28) and from a complete rat TAT cDNA clone (8). A sequence of 84 nucleotides about 600 bp from exon B was identified as the first exon of the human TAT gene by its sequence similarity to the 73 nucleotide exon A of rat TAT. To identify the exon/intron boundaries for intron 1, a cDNA clone obtained by reverse transcriptase−PCR using primers from exons A and B was sequenced (see Material and Methods). All exon/intron boundaries could thus be identified unequivocally by comparison of human TAT cDNA with genomic sequences, all conforming to the GT/AG splice junction rule (29).

The transcription start site of the human TAT gene was determined by primer extension analysis using an oligonucleotide complementary to nucleotides +49 to +65 in exon A as primer. A single transcription start site at nucleotide +1 was observed (not shown), which coincides with the major start site in the rat TAT gene (6, 8). The sequence of cDNA clone phcTAT3a-6 spans almost the entire exon L, but terminates 82 nucleotides 5′ to a polyadenylation signal sequence AATAAA found in the genomic sequence. As this polyadenylation sequence is conserved and functional in the rat, and as the surrounding sequences are highly conserved in the two species, we expect the 3′ end of exon L to be at position +2754, as indicated in Fig. 2. The human TAT gene thus spans about 10.9 kb. This compares with a size of 11 kb and 9.2 kb for the rat (6) and mouse (7) TAT genes, respectively.

The total size of the human TAT exons of 2754 bp as given in Fig. 2 is in good agreement with the size of the major TAT mRNA species of about 3000 nucleotides obtained from Northern blots (Fig. 3) considering the poly A tail. An open reading frame starting with the ATG codon at position +97 in exon B and terminating with the TAG stop codon at position +1459 in exon L codes for a 454 amino acid protein of molecular weight 50,399 dalton which is in reasonable agreement with the value of 95,500 dalton reported for the purified human TAT dimer (9). The sequence identity between the human and rat TAT mRNAs (counting deletions and insertions like single nucleotide substitutions) is 58.7% in the noncoding regions and 87.1% in the coding region. The sequence identity between the two TAT proteins is 91.6 %. This identity goes up to 95.2%, if one excludes the first 78 amino acid residues coded for in exon B, where 20 of the 38 amino acid substitutions are found. The clustering of amino acid changes in the amino terminal segment of the TAT protein is in keeping with the fact that the first 64 amino acid residues of rat TAT can be released by trypsin treatment without any loss of enzyme activity (30).

The 5′ flanking regions of the human and rat TAT genes show large stretches of sequence conservation (Fig. 2). In footprinting experiments, multiple protein binding sites have been identified in the rat TAT gene, covering most of the first 200 bp upstream of the cap site (31). This region shows about 64% conservation in the human sequence. A 41 nucleotide stretch from −595 to −635 showing 83% sequence identity includes the conserved hexanucleotide TGTTCT, the highly conserved 'right half' sequence of the 15 nucleotide GRE consensus (4). Five such sequences are found within the first 2900 bp in the 5′ flanking region of the rat TAT gene (U. Strähle, personal communication; 28). However, only GRE II and III at position −2500 are functional, while the other three, including GRE V, are not (4; see also below).

The human TAT gene is the second gene coding for a transaminase the exon/intron structure of which is known at the DNA sequence level, the first being the mouse gene for mitochondrial aspartate aminotransferase (AAT) (32). Marginally significant alignments of rat TAT with the cytosolic and mitochondrial isoenzymes of vertebrate AAT have been described yielding about 13% amino acid identity (30, 33). However, these alignments are not reflected in a conservation of exon/intron junctions in the genes for human TAT and mouse mitochondrial AAT.

Alternative poly A addition sites in the human TAT gene

In the 3′ untranslated region, there are two polyadenylation signal sequences AATAAA, at positions +2070 and +2738 (overlined in Fig. 2). Both are conserved in the rat, but only the 3′ signal is used in this species, giving rise to a mRNA of about 2400 bases (19; Fig. 3 C). Two mRNA species with sizes of about 3000 and 2300 bases are found, however, in human liver RNA hybridized with a TAT cDNA probe 5′ to position +2070 (probe A, Fig. 3 B, C). If a probe 3′ to position +2070 (and 3′ to the Alu element in exon L) is used, only the larger mRNA species is seen (probe B, Fig. 3 B, C). This indicates that the smaller transcript, which by densitometry of the Northern blot autoradiograph makes up about 10% of the amount of the larger transcript, could in fact result from use of the polyadenylation signal at +2070. This interpretation is corroborated by sequence analysis of cDNA clones (Fig. 3 A). While clone phcTAT2-16

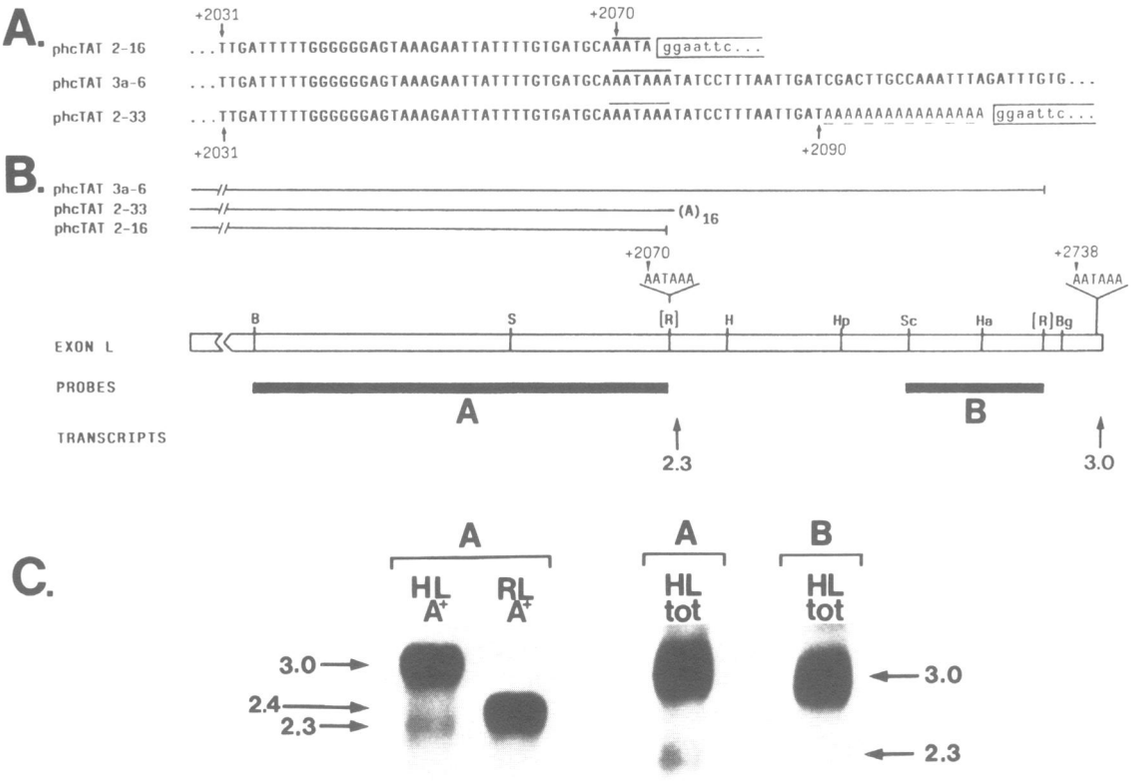


Figure 3: Two mRNA species from the human TAT gene. A:The 3′ sequences of cDNA clones phcTAT2-16 and phcTAT2-33 in comparison with the sequence of clone phcTAT3a-6. The polyadenylation signal at +2070 is overlined and the EcoRI cloning linkers are boxed. The poly A addition site at +2090 in clone phcTAT2-33 is indicated; the poly A stretch is underlined. B:Alignment of the 3′ ends of the cDNA clones shown in A with exon L. Restriction sites are designated as in legend to Fig. 1. The position of the two polyadenylation signals in exon L are indicated (see Fig. 2). Arrows indicate the poly A addition sites of the 2.3 kb and 3.0 kb transcripts. Probe A, 660 bp BamHI/EcoRI fragment from phcTAT2-16. Probe B, 300 bp ScaI/EcoRI fragment from phcTAT3a-6. C: Northern blot analysis of poly(A)$^+$ RNA (A$^+$) and total RNA (tot) from human (HL) and rat liver (RL). 5 μg of poly (A)$^+$ RNA and 50 μg of total RNA were separated on 1.2% agarose/formaldehyde gels, transferred to nylon filters and hybridized to probes A or B (shown in B). The size in kb of the rat TAT mRNA (19) and of the two human TAT transcripts are indicated, deduced from 28 S and 18 S ribosomal RNA run in parallel (not shown). Note that the two left and the two right lanes are from different experiments.

terminates at position +2073 within the polyadenylation signal sequence, another cDNA clone, phcTAT2-33, extends for some nucleotides downstream up to position +2090 and then terminates in a run of 16 A residues, not found in the cDNA sequence of phcTAT3a-6.

We note two sequences with significant homologies to sequence motifs frequently found 20 to 50 nucleotides 3′ to functional poly A signals. The sequence TTGCCAAATTTA, 20 nucleotides 3′ to the poly A signal at + 2070, resembling the 12bp motif TTG/ANNNTTTTTT (34), and the sequence TGTGTGCA at position + 2111, with similarity to the 8 bp consensus motif YGTGTTYY (35). Only very weak homologies to these motifs are found in the corresponding region 3′ of the non-functional poly A signal in the rat. Grange et al. (8) mention a weak Northern blot signal compatible with a rat TAT mRNA polyadenylated in relation to this poly A signal, but no data are shown. We, and others (3, 19) have not observed such a rat mRNA species. If it exists, it must amount to less then 1% of the 2.4 kb mRNA species.

Alu element insertion in the TAT 3′ exons in humans, apes and old world monkeys

Within the 3′ untranslated region of the human TAT gene, a complete Alu element is found at position +2131 to +2481 (boxed in Fig. 2), which shows 90% sequence identity to the Alu consensus sequence of Kariya et al. (36). It is inserted in opposite transcriptional orientation relative to the TAT gene and carries a typical A rich stretch at its 3′ end, which is interrupted by an unusual sequence of 18 nucleotides (CTTATATTGAA-CTTCTCT, written from the Alu 'sense' strand). None of the 41 Alu elements analyzed by Karya et al. (36) show such a long and atypical sequence within the 3′ A rich region. The Alu element is flanked by a 15 bp perfect direct repeat (indicated by arrows in Fig. 2). This repeat corresponds to a sequence found only once in the rat TAT cDNA, at position 2080 to 2094 (8), and deviates from this sequence at 3 positions. Therefore, the flanking direct repeats are the consequence of the Alu element insertion into human TAT exon L.

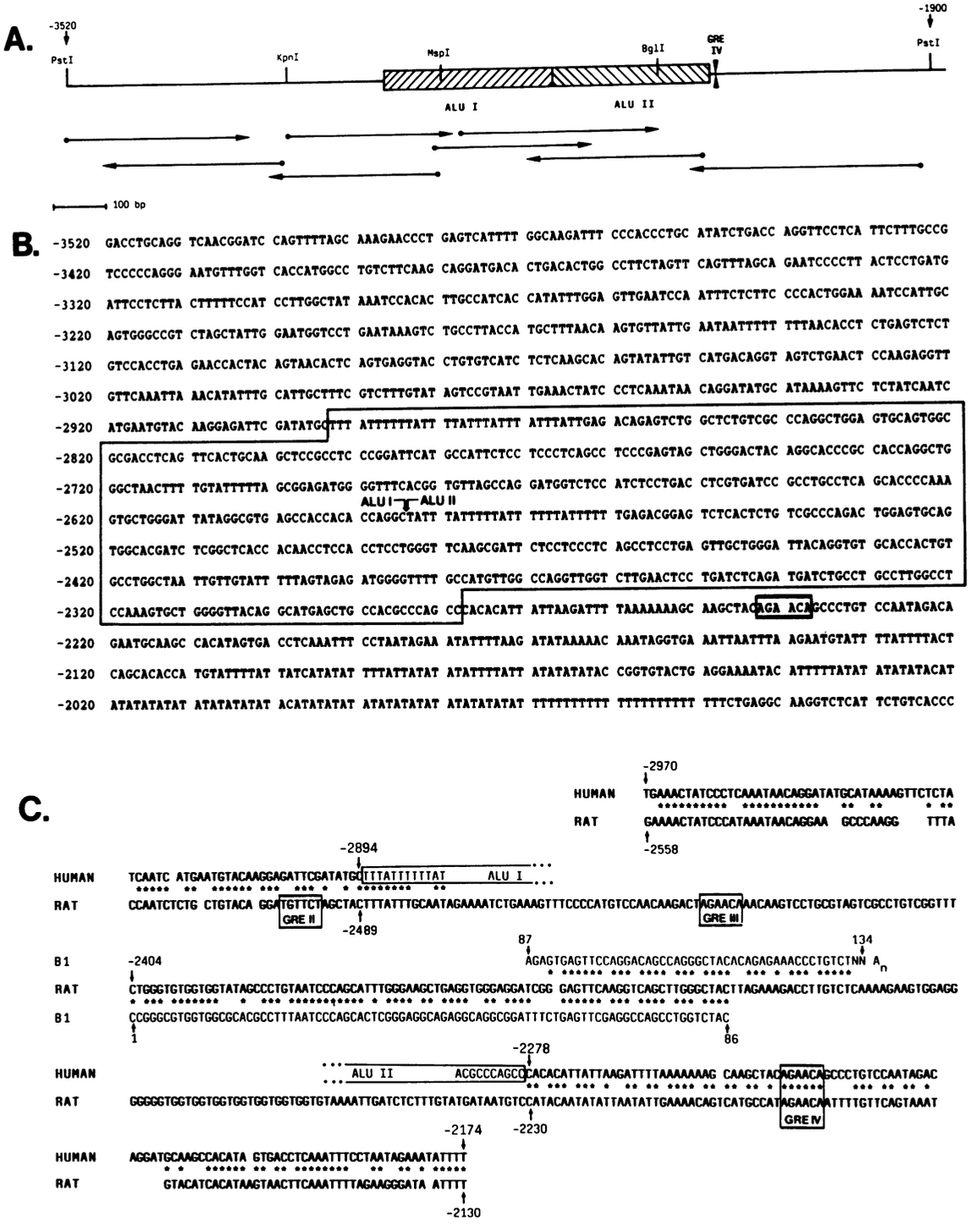


Figure 4: Sequence of the −3.5 kb to −1.9 kb region of the human TAT gene and comparison with sequences from the −2.5 kb rat GRE region. A:Restriction map of the 1.6 kb PstI fragment from the left end of the λhTAT1 insert (see Fig. 1). The righthand PstI site at positon −1900 is defined by restriction mapping, not by sequencing. Alu elements I and II are represented by hatched boxes, the position of the GRE IV motif is indicated. Arrows indicate direction and extent of DNA sequences determined. B:DNA sequence from position −3520 to −1921. The two Alu elements are boxed, as is the conserved GRE IV motif. C:The rat TAT sequence from position −2558 to −2130 (4, 28; U. Strähle, personal communication) in comparison with the human TAT sequence (position −2970 to −2174), and with the B1 mouse consensus sequence (38). Sequences are aligned to maximize for identities, indicated by asterisks, while empty spaces denote gaps. Only the beginning and end of the human Alu sequences are shown, while the rat sequence is given uninterrupted. The highly conserved hexanucleotide motif TGTTCT or AGAACA of the GRE consensus present in the rat GRE elements II, III and IV are boxed.

The fact that the sequences of the Alu flanking repeats are still identical, but differ in 3 out of 15 nucleotides from the rat sequence, could point to the Alu insertion as a more recent event in primate evolution. To address this question, Southern blot experiments were performed with different primate DNAs, using a probe 5′ to the Alu element (probe A from Fig. 3) and appropriate restriction enzyme cuts. These experiments demonstrated presence of the Alu element in DNA from human,

apes and gibbon, all hominoids, and also in the African green monkey, a cercopithecoid (data not shown; 27). Therefore, the Alu insertion into the TAT 3′ exon must have occured before the split of the cercopithecoid and hominoid lineages 25–34 milllion years ago (37). Results obtained for a new world monkey and a lemur were not conclusive.

Alu insertions interrupt 5′ flanking sequences with similarities to the −2.5 kb rat GRE region

Part of the glucocorticoid induction of the rat TAT gene is mediated by two GREs, GRE II and GRE III, located about 2500 bp from the start site of transcription (4). To search for conserved sequences, a 1.6 kb PstI fragment from the left end of the insert in λhTAT1 (Fig. 1) was subcloned, and its sequence determined according to the strategy shown in Fig. 4 A. A prominent feature in the resulting sequence, which spans positions −1921 to −3520 (Fig. 4 B), are two complete Alu elements arranged as a head to tail dimer, labelled Alu I and Alu II in Fig. 4. Their orientation relative to the TAT gene is the same as for the Alu element in exon L (Fig. 1 A). Both elements show the typical A rich sequence at their 3′ ends and have sequence indentities to the Alu consensus (36) of 86% (Alu I) and 88% (Alu II), but are not flanked by direct repeats.

A comparison of the sequence from Fig. 4 B with the sequences from positon −2100 to −2900 of the rat TAT gene in a dot matrix analysis (not shown) revealed only two short stretches of sequence identity just at the sequences flanking the Alu elements. A detailed analysis of the respective human and rat sequences is shown in Fig. 4 C. Beyond the sequence stretches shown, the sequence similarities break off completely. The region 5′ to Alu I shows similarities to rat GRE II and adjacent sequences, while the sequences 3′ to Alu II are clearly related to rat sequences around the non-functional GRE IV element, with complete conservation of the highly conserved hexanucleotide AGAACA of the 15 nucleotide GRE consensus sequence (4). The corresponding sequence TGTTCT of the functional rat GRE II is mutated to GATTCG in the human sequence, while the motif TGTACA, the less conserved 'left half' hexanucleotide of the GRE consensus, is unchanged. However, the two hexanucleotide motifs are separated by four instead of three nucleotides by an additional A residue (human position −2909). Due to these sequence changes, it is highly unlikely that GRE II can still function in the human. The GRE III element is completely missing from the human sequence and is replaced by the Alu elements. As might be expected, the 1.6 kb Pst I fragment does not confer glucocorticoid inducibility on a reporter gene in gene transfer experiments (C. Reinhard and G. Scherer, unpublished data). The conserved GRE V element at position −620 (Fig. 2) is likewise inactive. The regulatory elements responsible for the observed steroid induction of human TAT (10, 11) must reside elsewhere, possibly further upstream as recently demonstrated for an additional regulatory region at −5.4 kb in the rat TAT gene (5).

The presence of Alu elements within this region of the human TAT gene lead us to search for the Alu related B1 element in the corresponding region of the rat TAT gene. Such an element, not noticed previously, is found 3′ to rat GRE III, starting at position −2404. The B1 element is not complete, and can best be aligned with the 134 nucleotide B1 mouse consensus sequence (38) in the way shown in Fig. 4 C. It terminates in a very short A stretch, and is not flanked by tandem repeats. The finding of a rat B1 element at a position corresponding to the two human Alu elements is intriguing; however, as these elements are inserted in opposite orientation in their respective host sequences, a causal evolutionary relationship between them is not easily envisaged.

ACKNOWLEDGEMENTS

We thank E.-M. Westphal for alerting us to the presence of the Alu III element early in this study, W.Schempp for a gift of primate DNAs, C.Ebensperger and R.Jäger for help with PCR analysis, and U.Strähle for allowing us to use his rat TAT DNA sequences. We are grateful to B.Darimont, R.Jäger und W.Schmid for helpful comments on the manuscript, and E.Kunstmann for her excellent secretarial assistance. This work was supported by a grant from the Deutsche Forschungsgemeinschaft (Sche 194/4-3).

REFERENCES

1. Granner, D.K. and Hargrove, J.L. (1983) Mol. Cell. Biochem. **53/54**, 113–128.
2. Hashimoto, S., Schmid, W. and Schütz, G. (1984) Proc. Natl. Acad. Sci. USA **81**, 6637–6641.
3. Schmid, E., Schmid, W., Jantzen, M., Mayer, D., Jastorff, B. and Schütz, G. (1987) Eur. J. Biochem. **165**, 499–506.
4. Jantzen, H.-M., Strähle, U., Gloss, B., Stewart, F., Schmid, W., Boshart, M., Miksicek, R. and Schütz, G. (1987) Cell **49**, 29–38.
5. Grange, T., Roux, J., Rigaud, G. and Pictet, R. (1989) Nucl. Acids Res. **17**, 8695–8709.
6. Shinomiya, T., Scherer, G., Schmid, W., Zentgraf, H. and Schütz, G. (1984) Proc. Natl. Acad. Sci. USA **81**, 1346–1350.
7. Müller, G., Scherer, G., Zentgraf, H., Ruppert, S., Herrmann, B., Lehrach, H. and Schütz, G. (1985) J. Mol. Biol. **184**, 367–373.
8. Grange, T., Guénet, C., Dietrich, J.B., Chasserot, S., Fromont, M., Befort, N., Jami, J., Beck, G. and Pictet, R. (1985) J. Mol. Biol. **184**, 347–350.
9. Andersson, S.M. and Pispa, J.P. (1982) Clin. Chim. Acta **125**, 117–123.
10. Räihä, N.C.R. and Schwartz, A.L. (1973) Enzyme **15**, 330–339.
11. Andersson, S.M. (1981) Early Human Development **6**, 165–169.
12. Buist, N.R.M., Kennaway, N.G. and Fellmann, J.H. (1985) In Bickel, H. and Wachtel, U. (eds.) Inherited diseases of amino acid metabolism, Thieme, Stuttgart New York, pp. 203–235.
13. Natt, E., Kao, F.-T., Rettenmeier, R. and Scherer, G. (1986) Hum. Genet. **72**, 225–228.
14. Westphal, E.-M., Burmeister, M., Wienker, T.F., Lehrach, H., Bender, K. and Scherer, G. (1987) Genomics **1**, 313–319.
15. Westphal, E.-M., Natt, E., Grimm, T., Odievre, M. and Scherer, G. (1988) Hum. Genet. **79**, 260–264.
16. Natt, E., Westphal, E.-M., Toth-Fejel, S.E., Magenis, R.E., Buist, N.R.M., Rettenmeier, R. and Scherer, G. (1987) Hum. Genet. **77**, 352–358.
17. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning. A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Habor, NY.
18. Benton, W.D. and Davis, R.W. (1977) Science **196**, 180–182.
19. Scherer, G., Schmid, W., Strange, C.M., Röwekamp, W. and Schütz, G. (1982) Proc. Natl. Acad. Sci. USA **79**, 7205–7208
20. Chandra, T., Stackhouse, R., Kidd, V.J. and Woo, S.L.C. (1983) Proc. Natl. Acad. Sci. USA **80**, 1845–1848.
21. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA **74**, 5463–5467.
22. Maxam, A.M. and Gilbert, W. (1980) Methods Enzymol. **65**, 499–560.
23. Kawasaki, E.S. and Wang, A.M. (1989) In Erlich, H.A. (ed.), PCR Technology–Principles and Application for DNA Amplification, Stockton Press, New York, pp. 89–97.
24. Auffray, C. and Rougeon, F. (1980) Eur. J. Biochem. **107**, 303–314.
25. Feinberg, A.P. and Vogelstein, B. (1983) Anal. Biochem. **132**, 6–13.
26. Peters, C.W.B., Kruse, U., Pollwein, R., Grzeschik, K.-H. and Sippel, A.E. (1989) Eur. J. Biochem. **182**, 507–516.
27. Rettenmeier, R. (1988) Ph. D. thesis, Albert-Ludwigs-Universität, Freiburg.
28. Oddos, J., Grange, T., Carr, K.D., Matthews, B., Roux, J., Richard-Foy, H. and Pictet, R. (1989) Nucl. Acids Res. **17**, 8877–8878.
29. Breathnach, R. and Chambon, P. (1981) Ann. Rev. Biochem. **50**, 349–383.

30. Hargrove, J.L., Scoble, H.A., Mathews, W.R., Baumstark, B.R. and Biemann, K. (1989) J. Biol. Chem. **264**, 45–53.
31. Becker, P.B., Ruppert, S. and Schütz, G. (1987) Cell **51**, 435–443.
32. Tsuzuki, T., Obaru, K., Setoyama, C. and Shimada, K. (1987) J. Mol. Biol. **198**, 21–31.
33. Mehta, P.K., Hale, T.I. and Christen, P. (1989) Eur. J. Biochem. **186**, 249–253.
34. Renan, M.J. (1987) Gene **60**, 245–254.
35. McLauchlan, J., Gaffney, D., Whitton, J.L. and Clements, J.B. (1985) Nucl. Acids Res. **13**, 1347–1368.
36. Kariya, Y., Kato, K. Hayashizaki, Y., Himeno, S., Tarui, S. and Matsubara, K. (1987) Gene **53**, 1–10.
37. Sibley, C.G. and Ahlquist, J.E. (1987) J. Mol. Evol. **26**, 99–121.
38. Krayev, A.S., Kramerov, D.A., Skryabin, K.G., Ryskow, A.P., Bayev, A.A. and Georgiev, G.P. (1980) Nucl. Acids Res. **8**, 1201–1215.