Biochem. J. (1991) **278**, 809–816 (Printed in Great Britain)

Evidence that gene G7a in the human major histocompatibility complex encodes valyl-tRNA synthetase

Shie-Liang HSIEH and R. Duncan CAMPBELL*
M.R.C. Immunochemistry Unit, Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, U.K.

At least 36 genes have now been located in a 680 kb segment of DNA between the class I and class II multigene families within the class III region of the human major histocompatibility complex on chromosome 6p21.3. The complete nucleotide sequence of the 4.3 kb mRNA of one of these genes, G7a (or BAT6), has been determined from cDNA and genomic clones. The single-copy G7a gene encodes a 1265-amino-acid protein of molecular mass 140457 Da. Comparison of the derived amino acid sequence of the G7a protein with the National Biomedical Research Foundation protein databases revealed 42% identity in a 250-amino-acid overlap with *Bacillus stearothermophilus* valyl-tRNA synthetase, 38.0% identity in a 993-amino-acid overlap with *Escherichia coli* valyl-tRNA synthetase (*val RS*), and 48.3% identity in a 1043-amino-acid overlap with *Saccharomyces cerevisiae* valyl-tRNA synthetase. The protein sequence of G7a contains two short consensus sequences, His-Ile-Gly-His and Lys-Met-Ser-Lys-Ser, which is the typical signature structure of class I tRNA synthetases and indicative of the presence of the Rossman fold. In addition, the molecular mass of the G7a protein is the same as that of other mammalian valyl-tRNA synthetases. These features and the high sequence identity with yeast valyl-tRNA synthetase strongly support the fact that the G7a gene, located within the major histocompatibility complex, encodes the human valyl-tRNA synthetase.

INTRODUCTION

The human major histocompatibility complex (MHC) occupies a segment of approximately 4 centimorgans (cM) on the short arm of chromosome 6 and contains the highly polymorphic class I (HLA-A, -B and -C) and class II (DR, DQ, DP) genes. These genes are responsible for coding polymorphic cell surface proteins involved in the presentation and recognition of foreign antigens during immune responses (Strachan, 1987; Trowsdale, 1987; Davis & Bjorkman, 1988). In man, these two gene clusters are separated by ~ 1100 kb of DNA termed the class III region (Carroll *et al.*, 1987; Dunham *et al.*, 1987), which contains a number of unrelated genes, including those encoding the complement proteins C2, Factor B and C4 (Carroll *et al.*, 1984), the enzyme cytochrome *P*-450 steroid 21-hydroxylase (*CYP21*) (Carroll *et al.*, 1985; White *et al.*, 1985), the cytokines tumour necrosis factors α and β (Spies *et al.*, 1986; Carroll *et al.*, 1987; Dunham *et al.*, 1987) and three genes encoding members of the major heat-shock protein HSP70 family (Sargent *et al.*, 1989; Milner & Campbell, 1990).

The whole of the class III region has now been cloned in overlapping cosmid and yeast artificial chromosome clones (Sargent *et al.*, 1989; Spies *et al.*, 1989; Kendall *et al.*, 1990; Ragoussis *et al.*, 1991). A detailed characterization of the cloned DNA has led to the discovery of at least 25 novel genes in this region. A large number of these novel genes appear to be associated with HTF (*Hpa*II tiny fragment)-islands, CpG-rich sequences that are invariably found at the 5′ ends of genes that are ubiquitously expressed (Bird, 1986, 1987; Gardiner-Garden & Frommer, 1987). One of these HTF-islands lying within 10 kb of the *HSP70* genes in the class III region is associated with a gene, labelled G7a (Dunham *et al.*, 1990) or BAT6 (Spies *et al.*, 1989), that encodes a ~ 4 kb mRNA which is expressed in all cell types so far analysed.

The present paper reports the cloning and sequence analysis of the G7a cDNA. Comparison of the derived amino acid sequence of G7a with those in the protein databases revealed the strongest identity with *Saccharomyces cerevisiae* valyl-tRNA synthetase (525 identities out of 1095 possible matches). In addition, there is a domain in the G7a protein which is absent in the valyl-tRNA synthetase of lower eukaryotes and prokaryotes and which shows strong identity with brine shrimp (*Artemia salina*) elongation factor 1 γ-chain. Many of the features of the G7a protein deduced from the amino acid sequence are in accordance with the biochemical characteristics of mammalian valyl-tRNA synthetases based on studies of the rabbit enzyme (Bec *et al.*, 1989). Thus it is extremely likely that the human valyl-tRNA synthetase is encoded by the G7a gene in the class III region of the MHC.

MATERIALS AND METHODS

Cloning and nucleotide sequence analysis

For shotgun sequencing, the 3.6 kb cDNA insert of pG7a-1 was purified on a low-gelling-temperature (LGT) agarose gel, then digested with *Hin*fI, *Dde*I, *Msp*I and *Ava*II, and the fragments were ligated to M13mp10. To overlap contigs, *Sma*I, *Nco*I and *Bam*HI fragments from the 3.6 kb cDNA insert were also cloned and sequenced. All nucleotide sequencing was carried out by the dideoxy chain termination method (Sanger *et al.*, 1979). Single-stranded DNA was recovered from the pBluescript or M13 subclones and sequenced using the Sequenase system (U.S. Biochemicals, Cleveland, OH, U.S.A.). In the case of the pBluescript and pGEM 3Zf clones, the helper phage M13K07 was used for single-stranded DNA recovery, in the presence of kanamycin. Nucleotide sequences were obtained using the M13 universal primer (GTAAAACGACGGCCAGT, starting from the −21 position of the M13 phage), and both dGTP/ddGTP

Abbreviations used: MHC, major histocompatibility complex; HTF, *Hpa*II tiny fragment; SSC, 0.15 M-NaCl/0.015 M-sodium citrate; PMA, phorbol 12-myristate 13-acetate; LGT, low-gelling temperature.

* To whom all correspondence should be addressed.

The nucleotide sequence data reported will appear in the EMBL, Genbank and DDBJ Nucleotide Sequence Databases under the accession number X59303.

and dITP/ddITP (where dd is dideoxy) were used for sequencing reactions.

Southern blot analysis of cosmid DNA and genomic DNA

Both cosmid DNA (1 μg) and genomic DNA (5 μg) were digested with the appropriate restriction enzyme under the conditions recommended by the supplier. The digested DNA was fractionated on a 0.8 % agarose gel, transferred to nitrocellulose paper (Southern, 1975), and hybridized with ^{32}P-labelled probes. Probes were labelled directly in LGT agarose (FMC Products) by random hexanucleotide priming (Feinberg & Vogelstein, 1984). Blots were hybridized for 24 h at 42 °C in 50 % formamide/5 × Denhardt's solution/10 % dextran sulphate/1 M-NaCl/50 mM-Tris/HCl (pH 7.4)/0.1 % SDS containing 100 μg of sonicated salmon sperm DNA/ml. High-stringency washes were performed at 65 °C in 0.1 × standard sodium citrate (SSC)/0.2 % SDS for 1 h. Blots were autoradiographed between two intensifying screens at −70 °C for 1–5 days.

Isolation of RNA and Northern blot analysis

The cell lines U937, activated U937, HepG2, Raji, Molt4, HeLa and HL-60 were grown in tissue culture to densities of (1–2) × 10^6 cells/ml. The activated U937 cells were stimulated using PMA (phorbol 12-myristate 13-acetate; Sigma) for 3 days prior to their collection. Total RNA was extracted by the guanidinium isothiocyanate lysis method and caesium chloride ultracentrifugation (Chirgwin *et al.*, 1979; Maniatis *et al.*, 1982). Samples of total RNA (15 μg) were fractionated in 0.8 % agarose/formaldehyde denaturing gels and transferred on to nitrocellulose (Fourney *et al.*, 1988). Northern blots were hybridized with ^{32}P-labelled probes under the conditions described for genomic Southern blots, as outlined above. High-stringency washing was carried out at 65 °C in 0.2 × SSC/0.1 % SDS for 1 h followed by autoradiography at −70 °C between intensifying screens for 2 days.

RNAase protection

Transcription mapping of the 5′ end of the G7a gene was essentially by the method of Melton *et al.* (1984). A 2.2 kb genomic fragment (see Fig. 1) was subcloned into the pGEM 3Zf vector (Promega) and then linearized using the restriction endonuclease *Bss*HII. A riboprobe was derived by using either Sp6 polymerase or T7 polymerase for transcription. RQ1 RNAase-free DNAase was then added to a concentration of 1 unit/μg following transcription to remove the DNA template. A 40–100 μg sample of total RNA was mixed with 1 × 10^6 c.p.m. of the riboprobe, in a final volume of 30 μl, in 80 % formamide/40 mM-Pipes (pH 6.7)/0.4 M-NaCl/1 mM-EDTA, denatured by boiling for 5 min and incubated overnight at 45 °C. After hybridization, 300 μl of 300 mM-NaCl/5 mM-EDTA/10 mM-Tris/HCl (pH 7.5) was added with RNAase A (50 μg/ml)/RNAase T1 (6 μg/ml), and the mixture was incubated at 37 °C for 30 min. SDS (20 μl of 10 % solution) and 10 μl of freshly prepared protease K (10 mg/ml in water) were then added to the mixture and incubated at 37 °C for 30 min. After phenol/chloroform (1:1, v/v) extraction, the [^{32}P]RNA hybrids were ethanol-precipitated with the addition of calf liver tRNA as carrier. The products were analysed by electrophoresis on 6 % polyacrylamide/7 M-urea gels.

Computer analysis of cDNA and amino acid sequences

The sequence analyses and comparisons were carried out using the programs of Staden (1986) on the Oxford University Computing System. The PIR Protein and PIR Protein New databases were searched using DBSEARCH in order to find sequences similar to the G7a amino acid sequence. Programs NIP and SAP were used for the analysis of nucleotide sequence and amino acid sequence respectively.

RESULTS

Isolation and characterization of cDNA clones

A 9 kb *Bgl*II/*Xho*I fragment from the cosmid cos10S (Fig. 1*a*), which contains a potential HTF-island (Sargent *et al.*, 1989), was radiolabelled and used to probe a U937 cDNA library (Simmons & Seed, 1988) constructed in the CDM 8 vector (Seed, 1987). Fifteen positive clones were characterized, and the longest cDNA insert obtained (3.6 kb) was subcloned into the *Xba*I site of pBluescript KS$^+$ (pG7a-1). The sequence of the insert was determined by a combination of shotgun cloning and specific fragment cloning, as described in the Materials and methods section. The complete sequence of both strands was obtained with a degeneracy of about 4. Since the 3.6 kb cDNA insert of pG7a-1, which includes a poly(A) tail, contains only part of the 5′ end sequence (Fig. 1*b*), a 310 bp *Xho*I/*Eco*RI fragment derived from the extreme 5′ end of the clone was used to identify those clones which extended pG7a-1 at the 5′ end. DNA from all of the positive clones was double-digested with *Xho*I/*Eco*RI, blotted on to nitrocellulose and hybridized with this probe. One of the positive clones, with an insert of ~ 2.1 kb, which contains an 800 bp *Xho*I/*Eco*RI fragment (pG7a-2) (Fig. 1*b*), was characterized further by DNA sequence analysis. From these two inserts, the full cDNA sequence was determined (see Fig. 4). This starts from the 5′ untranslated region and extends to a 18 bp poly(A) tail. The size of the mRNA deduced from cDNA sequence analysis and 5′ end mapping (Figs. 2*a* and 2*b*) was 4017 bp. This is very close to the size of 4.3 kb for the G7a mRNA estimated by Northern blot analysis (Fig. 3), assuming an average poly(A) tail length of ~ 150 bases.

RNAase protection and 5′ end mapping

Hybridization of the cDNA probes to Southern blots of restriction digests of cos10S DNA revealed that the G7a gene spans ~ 21 kb (Fig. 1*a*). In order to map the 5′ end of the gene, a 2.2 kb *Pvu*II fragment which contains the potential HTF-island (Fig. 1*a*) and which hybridized to the 5′ cDNA probe was subcloned into the vector pGEM 3Zf. The genomic insert was mapped with a number of different enzymes, including *Sac*II. Since the distance between the *Pvu*II and *Sac*II sites is 414 nucleotides in the cDNA sequence, but is more than 750 nucleotides in the genomic fragment, an intron must exist between these two restriction enzyme cutting sites. Sequence analysis of an 800 bp *Pvu*II/*Bss*HII genomic fragment confirmed this and revealed that an ~ 340 bp intron separates exons 1 and 2 (Figs. 1*b* and 2*b*). Radioactive RNA transcripts containing sequences complementary to the first exon and part of the second exon of the G7a gene were generated as described in the Materials and methods section. From the RNAase protection experiment, two protected fragments were found (Fig. 2*a*). The fragment of 271 bases corresponds to the 267 base fragment expected from protection of sequences in the second exon, extending from the *Pvu*II site to the 5′ nucleotide of this exon. The difference in size of 4 bases from that expected is most likely due to the slightly different migration velocities of DNA and RNA on polyacrylamide gels. The other protected fragment of 190 bases is derived from the first exon. By comparison of the cDNA and genomic sequences, this allowed the 5′ nucleotide to be assigned (Fig. 2*b*).

In the sequences up to −63 bp from the putative CAP site, there were no matches to the TATA or CAAT consensus sequences in the expected positions (Fig. 2*b*). However, there is

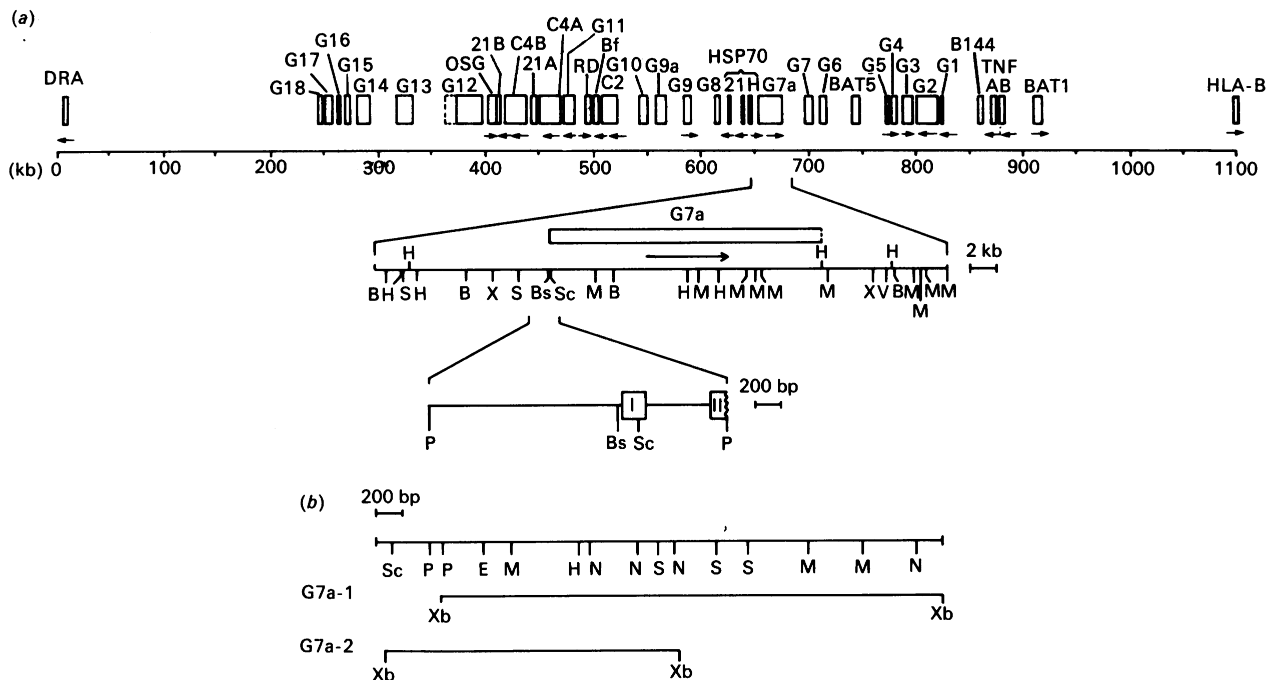

Fig. 1. Restriction maps of genomic and cDNA clones

(*a*) Location of the G7a gene in the MHC class III region. The direction of transcription of the genes, which are illustrated by open boxes, is indicated by arrows. The G7a gene has also been designated BAT6 (Spies *et al.*, 1989). Details of the molecular map can be found in Sargent *et al.* (1989), Spies *et al.* (1989) and Kendall *et al.* (1990). The expanded region is a restriction map of the cosmid cos10S (Sargent *et al.*, 1989). The 5′ end of the G7a gene is associated with an HTF-island defined by the presence of clustered sites for the rarely-cutting enzymes *Bss*HII and *Sac*II. The 3′ limit of the G7a gene in the genomic insert has not been accurately defined. Also shown is the 2.2 kb *PvuII* fragment which contains the 5′ region of the G7a gene, including exon 1 and part of exon 2. (*b*) Restriction map of the cDNA clones pG7a-1 and pG7a-2. The *XbaI* sites at the ends of the inserts are from the vector CDM8. The vector also contains unique *Xho*I sites that lie close to the *Xba*I sites (not shown). The abbreviations used for restriction enzymes are: B, *Bgl*II; Bs, *Bss*HII; E, *Eco*RI; H, *Hin*dIII; M, *Bam*HI; N, *Nco*I; P, *Pvu*II; S, *Sma*I; Sc, *Sac*II; X, *Xho*I; Xb, *Xba*I.

a GC-rich box containing two GGCGGG motifs at positions −34 and −50, which is the central component of binding sites for the transcriptional factor Sp1. Increasing numbers of higher-eukaryotic promoters are being found which lack a TATA box, but instead have a GC-rich region, such as the promoters of the genes encoding hypoxanthine phosphoribosyltransferase, dihydrofolate reductase, phosphoglycerate kinase and adenosine deaminase (Tsui & Siminovitch, 1987). These may be a feature of 'housekeeping genes', and G7a, which is probably also a housekeeping gene, also shares this feature. In addition, the 5′ region and exon 1 of the G7a gene contain a high percentage of CpG dinucleotides, a characteristic of HTF-island-associated genes (Bird, 1987).

Nucleotide sequence and derived amino acid sequence of G7a

Translation of the G7a cDNA sequence in three phases identified a single long open reading frame following the first Met codon at nucleotide 220 (Fig. 4). The DNA sequence around the proposed initiating Met complies well with the consensus sequence for vertebrate translation initiation sites (Kozak, 1984). The coding sequence is 3795 nucleotides in length and contains an unusually high percentage of guanidine and cytosine (G = 31.0 %, C = 30.8 %), and a relatively low percentage of adenosine and thymidine (A = 19.1 %, T = 19.1 %), which is remarkably different from most eukaryotic genes, in which the A+T percentage is higher than the G+C percentage. The reason for the high G+C content in G7a cDNA is due to both the high percentage of codons for Ala (9.4 %), Pro (7.9 %), Val (7.2 %), Glu (6.3 %) and Lys (4.5 %), and the codon usage bias of these five amino acids. The size of the 5′ and 3′ untranslated regions of the G7a mRNA are 219 bases and 72 bases respectively. In the 3′ untranslated region a canonical polyadenylation signal AATAAA lies 18 bases from the poly(A) tail (Fig. 4).

The open reading frame of G7a encodes a 1265-amino-acid polypeptide with a molecular mass of 140457 Da. Comparison with the NBRF protein databases revealed 42 % identity in a 250-amino-acid overlap with *Bacillus stearothermophilus* valyl-tRNA synthetase, 38.0 % identity in a 993-amino-acid overlap with *Escherichia coli* valyl-tRNA synthetase (*val RS*), and 48.3 % identity in a 1043-amino-acid overlap with *S. cerevisiae* valyl-tRNA synthetase (Fig. 5*a*). In addition, the unique *N*-terminal domain (amino acids 1–154) was found to display strong similarity (27.8 %) with the brine shrimp (*A. salina*) elongation factor 1 γ-chain (Fig. 5*b*). The molecular mass of the G7a gene product is the same as that of rabbit valyl-tRNA synthetase (Bec & Waller, 1989; Bec *et al.*, 1989), and bigger than the corresponding enzyme of lower eukaryotes and prokaryotes (125 kDa in *S. cerevisiae*, 108 kDa in *E. coli*) (Jordana *et al.*, 1987; Heck & Hatfield, 1988). There are six hydrophobic segments (amino acids 65–85, 99–119, 152–172, 336–356, 811–831 and 1161–1181) based on the method of Eisenberg *et al.* (1984) in G7a. Three of these segments (amino acids 152–172, 811–831 and 1161–1181) are associated with α-helix structure and could potentially serve as transmembrane domains. The theoretical isoelectric point (pI value) of the G7a protein is about 7.59. However, there is a strong basic segment in the *N*-terminal region (from amino acids 127 to 278), and the pI value is as great as 11 from amino acids 228 to 278, since this segment is rich in Lys and Arg residues. The protein sequence of G7a contains two short common consensus sequences, His-Ile-Gly-His and Lys-

Fig. 2. Mapping of the G7a transcription start point

(*a*) RNAase protection analysis using a 800 base riboprobe (lane 5) complementary to exon 1 and part of exon 2 of the G7a gene. The riboprobe was annealed to total RNA from U937 cells (lane 2), calf thymus tRNA (lane 3) and no RNA (lane 4), digested with RNAase A and T1, and the resulting protected fragments were fractionated in a 6% polyacrylamide/7 M-urea gel. The sizes of the protected fragments in lane 2 are shown on the right. The size markers on the left are taken from the M13mp10 sequencing ladder (lane 1). (*b*) Nucleotide sequence at the 5′ end of the G7a gene. The nucleotide assigned +1 on the basis of the RNAase protection analysis is indicated with an asterisk. The position of the intron was determined by comparison with the cDNA sequence.

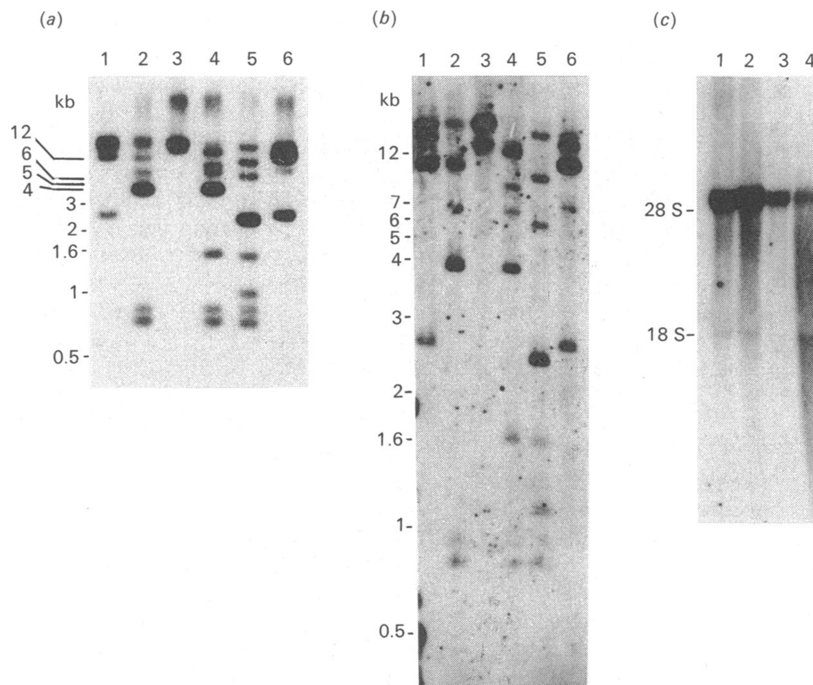

Fig. 3. Southern and Northern blot analysis

Cosmid (*a*) and genomic (*b*) Southern blot analysis using the 3.6 kb cDNA insert of pG7a-1. The cDNA insert was hybridized to Southern blots of cosmid cos10S and genomic DNA digested with *Hin*dIII (lane 1), *Bam*HI (lane 2), *Bgl*II (lane 3), *Bam*HI/*Bgl*II (lane 4), *Bam*HI/*Hin*dIII (lane 5) or *Bgl*II/*Hin*dIII (lane 6). The genomic DNA was prepared from the HLA homozygous consanguineous cell line (HLA type: A2 Cw7 B7 C2C BfS C4A3 C4BQO DR2) used to construct the cosmid library. Numbers indicate the positions of DNA markers in kb. (*c*) Northern blot analysis. The pG7a-1 cDNA insert was hybridized to a Northern blot containing ~ 20 μg of total RNA from the cell lines U937 (lane 1), U937 stimulated with PMA (lane 2), Hep G2 (lane 3) and HL-60 (lane 4). The positions of migration of 28 S and 18 S RNA are indicated.

```
   1  GTCACAAAGGGGGGACACGTGGGCGCCGGCTGCCGGGGCGGCGATCTTAGGGAACTAGGGTCACCTGGAGAGCCGCCCACCGTCTCTGCC
  91  CGCTCGACTCCTCCGCCCGGGCCGCTCGGCGGTCCAGCCGCGGCCGGCGCCTGGCTGTGAGGTGGATTCCCGGCCCAGTCTGACCATCTC
 181  CCTCCAGTTTTTCCACTTCGTTCGGACCTTCTCATAACTATGTCCACCCTCTACGTCTCCCCTCACCCAGATGCCTTCCCCAGCCTCCGA
   1                                          M  S  T  L  Y  V  S  P  H  P  D  A  F  P  S  L  R
 271  GCCCTCATAGCCGCTCGCTATGGGGAGGCTGGGGAGGGTCCCGGATGGGGAGGAGCCCACCCCCGCATCTGTCTCCAGCCACCCCCGACT
  18  A  L  I  A  A  R  Y  G  E  A  G  E  G  P  G  W  G  G  A  H  P  R  I  C  L  Q  P  P  P  T
 361  AGCAGGACTAGCTTTCCCCCACCCCGCCTGCCGGCCCTGGAGCAGGGGCCCGGTGGGCTCTGGGTGTGGGGGGCCACGGCTGTGGCCCAG
  48  S  R  T  S  F  P  P  P  R  L  P  A  L  E  Q  G  P  G  G  L  W  V  W  G  A  T  A  V  A  Q
 451  CTGCTGTGGCCAGCAGGCCTGGGGGGCCCAGGGGGCAGCCGGGCGGCTGTCCTTGTCCAACAGTGGGTCAGTTACGCCGACACGGAGTTA
  78  L  L  W  P  A  G  L  G  G  P  G  G  S  R  A  A  V  L  V  Q  Q  W  V  S  Y  A  D  T  E  L
 541  ATACCAGCTGCCTGTGGAGCAACGCTGCCGGCCCTGGGACTCCGAAGCTCGGCCCAGGACCCCCAGGCTGTGCTGGGGGCCCTGGGCAGG
 108  I  P  A  A  C  G  A  T  L  P  A  L  G  L  R  S  S  A  Q  D  P  Q  A  V  L  G  A  L  G  R
 631  GCCCTGAGCCCCTTGGAGGAGTGGCTTCGGCTGCACACCTACTTGGCCGGGGAGGCCCCCACTCTGGCTGACCTGGCGGCTGTCACAGCC
 138  A  L  S  P  L  E  E  W  L  R  L  H  T  Y  L  A  G  E  A  P  T  L  A  D  L  A  A  V  T  A
 721  TTGCTGCTGCCTTTCCGATACGTCCTAGACCCACCTGCCCGCCGGATCTGGAATAATGTGACTCGCTGGTTTGTCACGTGTGTCCGACAG
 168  L  L  L  P  F  R  Y  V  L  D  P  P  A  R  R  I  W  N  N  V  T  R  W  F  V  T  C  V  R  Q
 811  CCAGAATTCCGAGCCGTGCTAGGAGAAGTGGTTCTATACTCAGGAGCCAGGCCTCTCTCTCATCAGCCAGGCCCCGAGGCTCCTGCCCTC
 198  P  E  F  R  A  V  L  G  E  V  V  L  Y  S  G  A  R  P  L  S  H  Q  P  G  P  E  A  P  A  L
 901  CCAAAGACAGCTGCTCAGCTCAAGAAAGAGGCAAAGAAACGGGAGAAGCTAGAGAAATTCCAACAGAAGCAGAAGATCCAACAGCAGCAG
 228  P  K  T  A  A  Q  L  K  K  E  A  K  K  R  E  K  L  E  K  F  Q  Q  K  Q  K  I  Q  Q  Q  Q
 991  CCACCTCCAGGGGAGAAGAAACCAAAACCAGAGAAGAGGGAGAAACGGGATCCTGGGGTCATTACCTATGACCTCCCAACCCCACCCGGG
 258  P  P  P  G  E  K  K  P  K  P  E  K  R  E  K  R  D  P  G  V  I  T  Y  D  L  P  T  P  P  G
1081  GAAAAGAAAGATGTCAGTGGCCCCATGCCCGACTCCTACAGCCCTCGGTATGTGGAGGCTGCCTGGTACCCTTGGTGGGAGCAGCAGGGC
 288  E  K  K  D  V  S  G  P  M  P  D  S  Y  S  P  R  Y  V  E  A  A  W  Y  P  W  W  E  Q  Q  G
1171  TTCTTCAAGCCAGAGTATGGGCGTCCTAATGTGTCAGCAGCAAATCCCCGAGGTGTCTTCATGATGTGCATCCCACCCCCCAATGTGACA
 318  F  F  K  P  E  Y  G  R  P  N  V  S  A  A  N  P  R  G  V  F  M  M  C  I  P  P  P  N  V  T
1261  GGCTCCCTGCACCTGGGCCATGCACTCACCAACGCCATCCAGGACTCCCTGACTCGATGGCACCGCATGCGTGGGGAGACCACCCTGTGG
 348  G  S  L  H  L  G  H  A  L  T  N  A  I  Q  D  S  L  T  R  W  H  R  M  R  G  E  T  T  L  W
1351  AACCCTGGCTGTGACCATGCAGGTATTGCCACCCAGGTGGTGGTGGAGAAGAAGCTATGGCGTGAGCAGGGACTGAGCCGGCACCAGCTG
 378  N  P  G  C  D  H  A  G  I  A  T  Q  V  V  V  E  K  K  L  W  R  E  Q  G  L  S  R  H  Q  L
1441  GGCCGCGAGGCCTTTCTACAGGAAGTCTGGAAGTGGAAGGAGGAGAAAGGTGACCGGATTTACCACCAGTTGAAGAAGCTTGGCAGCTCC
 408  G  R  E  A  F  L  Q  E  V  W  K  W  K  E  E  K  G  D  R  I  Y  H  Q  L  K  K  L  G  S  S
1531  TTGGACTGGGATCGAGCCTGTTTCACCATGGACCCTAAACTCTCAGCAGCTGTGACAGAGGCCTTTGTCCGGCTTCACGAGGAAGGCATC
 438  L  D  W  D  R  A  C  F  T  M  D  P  K  L  S  A  A  V  T  E  A  F  V  R  L  H  E  E  G  I
1621  ATCTATCGCAGTACCCGCCTTGTTAACTGGTCCTGCACCCTCAACTCCGCCATCTCTGACATTGAGGTGGATAAGAAGGAGCTGACAGGT
 468  I  Y  R  S  T  R  L  V  N  W  S  C  T  L  N  S  A  I  S  D  I  E  V  D  K  K  E  L  T  G
1711  CGCACCCTGCTCTCCGTGCCTGGCTACAAGGAGAAGGTGGAGTTCGGGGTCCTCGTGTCCTTTGCCTATAAGGTCCAAGGCTCAGATAGC
 498  R  T  L  L  S  V  P  G  Y  K  E  K  V  E  F  G  V  L  V  S  F  A  Y  K  V  Q  G  S  D  S
1801  GACGAGGAGGTGGTGGTGGCAACAACTCGGATCGAGACAATGCTGGGAGATGTGGCTGTAGCTGTGCACCCCAAAGATACCAGATACCAG
 528  D  E  E  V  V  V  A  T  T  R  I  E  T  M  L  G  D  V  A  V  A  V  H  P  K  D  T  R  Y  Q
1891  CACCTGAAGGGGAAGAACGTGATCCACCCATTCCTGTCTCGGAGCCTTCCCATTGTCTTCGATGAATTTGTGGACATGGACTTTGGCACA
 558  H  L  K  G  K  N  V  I  H  P  F  L  S  R  S  L  P  I  V  F  D  E  F  V  D  M  D  F  G  T
1981  GGTGCTGGGAAGATCACCCCCGCACATGACCAAAATGACTATGAAGTTGGGCAGCGGCACGGGCTGGAGGCCATCAGCATCATGGACTCC
 588  G  A  G  K  I  T  P  A  H  D  Q  N  D  Y  E  V  G  Q  R  H  G  L  E  A  I  S  I  M  D  S
2071  CGGGGGGGCCCTCATCAATGTGCCTCCGCCTTTCCTGGGCCTGCCCAGGTTTTGAGGCCAGGAAAGCGGTGCCTGGTGGCGCTGAAGGAG
 618  R  G  G  P  H  Q  C  A  S  A  F  P  G  P  A  Q  V  L  R  P  G  K  R  C  L  V  A  L  K  E
2161  CGGGGACTGTTCCGTGGCATTGAGGACAACCCCATGGTGGTGCCACTTTGCAACCGGTCGAAGGACGTGGTAGAGCCTCTGCTGCGGCCG
 648  R  G  L  F  R  G  I  E  D  N  P  M  V  V  P  L  C  N  R  S  K  D  V  V  E  P  L  L  R  P
2251  CAGTGGTACGTTCGCTGCGGGGAGATGGCCCAGGCTGCCAGCGCCGCTGTGACTCGGGGTGACCTCCGCATCCTGCCTGAGGCCCATCAG
 678  Q  W  Y  V  R  C  G  E  M  A  Q  A  A  S  A  A  V  T  R  G  D  L  R  I  L  P  E  A  H  Q
2341  CGCACATGGCATGCCTGGATGGACAACATCCGGGAGTGGTGCATTTCCAGGCAGCTGTGGTGGGGCCATCGCATCCCAGCCTACTTTGTC
 708  R  T  W  H  A  W  M  D  N  I  R  E  W  C  I  S  R  Q  L  W  W  G  H  R  I  P  A  Y  F  V
2431  ACTGTCAGTGACCCAGCGGTGCCCCCTGGGGAGGACCCTGATGGGCGGTACTGGGTGAGTGGACGCAATGAGGCGGAGGCCCGGGAGAAG
 738  T  V  S  D  P  A  V  P  P  G  E  D  P  D  G  R  Y  W  V  S  G  R  N  E  A  E  A  R  E  K
2521  GCAGCCAAGGAGTTCGGAGTGTCCCCTGACAAGATCAGTCTCCAGCAAGATGAGGATGTATTGGATACCTGGTTCTTCTCTGGCCTCTTC
 768  A  A  K  E  F  G  V  S  P  D  K  I  S  L  Q  Q  D  E  D  V  L  D  T  W  F  F  S  G  L  F
2611  CCCTTATCCATTTTGGGCTGGCCCAACCAGTCAGAAGACCTGAGTGTGTTCTACCCCGGGACACTGCTGGAGACCGGTCATGACATCCTC
 798  P  L  S  I  L  G  W  P  N  Q  S  E  D  L  S  V  F  Y  P  G  T  L  L  E  T  G  H  D  I  L
2701  TTCTTCTGGGTGGCCCGGATGGTCATGCTGGGCCTGAAGCTCACGGGCAGGCTGCCCTTTAGAGAGGTCTACCTCCATGCCATCGTGCGA
 828  F  F  W  V  A  R  M  V  M  L  G  L  K  L  T  G  R  L  P  F  R  E  V  Y  L  H  A  I  V  R
2791  GATGCTCACGGCCGGAAGATGAGCAAGTCTCTAGGCAATGTCATCGATCCCCTGGACGTCATCTATGGAATCTCCCTGCAGGGCCTCCAC
 858  D  A  H  G  R  K  M  S  K  S  L  G  N  V  I  D  P  L  D  V  I  Y  G  I  S  L  Q  G  L  H
2881  AACCAGCTGCTGAACAGCAACCTGGATCCCAGCGAGGTGGAGAAGGCCAAAGAAGGGCAGAAAGCTGACTTCCCAGCGGGGATTCCTGAA
 888  N  Q  L  L  N  S  N  L  D  P  S  E  V  E  K  A  K  E  G  Q  K  A  D  F  P  A  G  I  P  E
2971  TGTGGCACCGATGCTCTCCGGTTTGGATTATGTGCCTACATGTCCCAGGGTCGTGACATCAACCTGGATGTGAACCGGATACTGGGTTAC
 918  C  G  T  D  A  L  R  F  G  L  C  A  Y  M  S  Q  G  R  D  I  N  L  D  V  N  R  I  L  G  Y
3061  CGCCACTTCTGCAACAAGCTCTGGAATGCCACCAAGTTTGCCCTTCGTGGCCTTGGGAAGGGTTTTGTGCCCTCACCCACCTCCCAGCCC
 948  R  H  F  C  N  K  L  W  N  A  T  K  F  A  L  R  G  L  G  K  G  F  V  P  S  P  T  S  Q  P
3151  GGAGGCCATGAGAGCCTGGTGGACCGCTGGATCCGCAGCCGCCTGACAGAGGCTGTGAGGCTCAGCAATCAAGGCTTCCAGGCCTACGAC
 978  G  G  H  E  S  L  V  D  R  W  I  R  S  R  L  T  E  A  V  R  L  S  N  Q  G  F  Q  A  Y  D
3241  TTCCCGGCCGTCACCACTGCCCAGTACAGCTTCTGGCTCTATGAGCTCTGTGATGTCTACTTGGAGTGCCTGAAACCTGTACTGAATGGG
1008  F  P  A  V  T  T  A  Q  Y  S  F  W  L  Y  E  L  C  D  V  Y  L  E  C  L  K  P  V  L  N  G
3331  GTGGACCAGGTGGCAGCTGAGTGTGCCCGCCAGACCCTGTACACTTGCCTGGACGTTGGCCTGCGGCTGCTCTCACCCTTCATGCCCTTC
1038  V  D  Q  V  A  A  E  C  A  R  Q  T  L  Y  T  C  L  D  V  G  L  R  L  L  S  P  F  M  P  F
3421  GTGACGGAGGAGCTGTTCCAGAGGCTGCCCCGGAGGATGCCGCAAGCTCCCCCTAGCCTCTGTGTTACCCCCTACCCGGAGCCCTCAGAG
1068  V  T  E  E  L  F  Q  R  L  P  R  R  M  P  Q  A  P  P  S  L  C  V  T  P  Y  P  E  P  S  E
3511  TGCTCCTGGAAGGACCCCGAGGCAGAAGCCGCCCTTGAGCTGGCGCTAAGCATCACGCGAGCCGTGCGCTCCCTGCGGGCCGACTACAAC
1098  C  S  W  K  D  P  E  A  E  A  A  L  E  L  A  L  S  I  T  R  A  V  R  S  L  R  A  D  Y  N
3601  CTCACCCGGATCCGGCCTGACTGTTTCCTGGAAGTGGCGGATGAGGCCACGGGCGCCCTGGCATCGGCGGTGTCGGGCTACGTGCAGGCC
1128  L  T  R  I  R  P  D  C  F  L  E  V  A  D  E  A  T  G  A  L  A  S  A  V  S  G  Y  V  Q  A
3691  CTGGCCAGCGCAGGTGTGGTGGCTGTTCTGGCCCTGGGGGCTCCCGCCCCCCAGGGTTGCGCTGTGGCTCTGGCTTCTGATCGCTGCTCC
1158  L  A  S  A  G  V  V  A  V  L  A  L  G  A  P  A  P  Q  G  C  A  V  A  L  A  S  D  R  C  S
3781  ATCCACCTGCAGCTTCAGGGGCTGGTGGACCCTGCACGGGAGCTGGGCAAGCTGCAAGCCAAGCGAGTTGAGGCCCAGCGGCAGGCCCAG
1188  I  H  L  Q  L  Q  G  L  V  D  P  A  R  E  L  G  K  L  Q  A  K  R  V  E  A  Q  R  Q  A  Q
3871  CGTCTGCGGGAACGCCGTGCTGCCTCGGGCTATCCTGTCAAGGTGCCGCTCGAAGTCCAGGAGGCAGATGAAGCCAAGCTCCAACAGACA
1218  R  L  R  E  R  R  A  A  S  G  Y  P  V  K  V  P  L  E  V  Q  E  A  D  E  A  K  L  Q  Q  T
3961  GAAGCAGAGCTCAGGAAGGTGGATGAGGCCATCGCCCTATTCCAGAAGATGCTGTGATCCACCACCCAGCTTCACCCCTCACCCCCAGCG
1248  E  A  E  L  R  K  V  D  E  A  I  A  L  F  Q  K  M  L
4051  GCTCACCATGGGGATGGCAGCAATAAAATATTTCCCACAAAAAAAAAAAAAAAAAA
```

Fig. 4. cDNA sequence and deduced amino acid sequence of G7a

The complete nucleotide sequence of the G7a mRNA from the transcription start site to the poly(A) tail is shown. The first translation start codon (ATG) is at nucleotide position 220. The derived amino acid sequence of G7a is given immediately below the nucleotide sequence. Dots above the nucleotide sequence indicate every tenth nucleotide.

Met-Ser-Lys-Ser, which is the typical signature structure of class I tRNA synthetases and indicative of the presence of the Rossman fold (Eriani *et al.*, 1990). In addition to the high similarity of the G7a gene product to yeast valyl-tRNA synthetase, all of the primary structural features also support the fact that the G7a gene is equivalent to the human valyl-tRNA synthetase gene.

Genomic and cosmid Southern blotting

In order to determine the copy number of the valyl-tRNA synthetase gene, both genomic DNA and cos10S DNA, which contains the entire valyl-tRNA synthetase gene, were digested with different restriction enzymes, and the patterns were compared after hybridization of the Southern blot with the cDNA probe (pG7a-1) (Fig. 3). No extra bands could be found in the genomic Southern blot compared with the cos10S Southern blot, suggesting that the human valyl-tRNA synthetase gene is a single-copy sequence in the genome.

DISCUSSION

From the study of *E. coli* alanyl-tRNA synthetase, it is clear that functional domains of the enzyme are arranged in an approximately linear fashion along the sequence (Schimmel, 1987, 1989), so it is likely that the increasing length of the *N*-terminal domain from lower eukaryotes to mammals reflects the fact that each newly acquired domain has its own function which is independent of the common catalytic features. Deletion of the *N*-terminal domains does not affect the catalytic activity of alanyl-tRNA synthetase. It has been speculated that unique *N*-terminal domains, which have been labelled 'dispensible sequences', in different tRNA synthetases could serve functions other than catalysis, such as a role in regulation of transcription or translation, or RNA splicing (Schimmel & Soll, 1979; Schimmel, 1987).

The strong sequence similarity of the unique *N*-terminal domain of human valyl-tRNA synthetase and brine shrimp (*A. salina*) elongation factor 1 γ-chain gives some indication about its function. Unlike the valyl-tRNA synthetases from prokaryotes and lower eukaryotes, which are monomeric enzymes of 110 kDa and 125 kDa respectively, that from various mammalian cell lines occurs as a high-molecular-mass entity (Ussery *et al.*, 1977; Kellermann *et al.*, 1982). Bec and co-workers (Bec & Waller, 1989; Bec *et al.*, 1989) have studied rabbit valyl-tRNA synthetase and have found that it always co-purifies with elongation factor 1H. Both rabbit and yeast valyl-tRNA synthetase display strong affinity toward the polyanionic support heparin–Ultrogel, a property not manifested by the corresponding prokaryotic enzyme. However, unlike the yeast enzyme, that of mammalian origin additionally exhibited hydrophobic properties. Based on these findings, these workers proposed that the mammalian valyl-tRNA synthetase has conserved the polycationic *N*-terminal domain that distinguishes the corresponding lower-eukaryotic enzyme from its prokaryotic counterpart, while acquiring a hydrophobic domain most likely responsible for its association to elongation factor 1H. This proposed structure fits exactly with the G7a amino acid sequence and the alignment of G7a with both *S. cerevisiae* valyl-tRNA synthetase and brine shrimp elongation factor 1 γ-chain. From the pairwise comparisons (Figs. 5*a* and 5*b*), the best alignment between G7a and the *S. cerevisiae* valyl-tRNA synthetase is from amino acid 1 of the yeast enzyme and amino acid 155 of G7a, and the best alignment between G7a and the *E. coli* enzyme is from the first Met of *E. coli* valyl-tRNA synthetase and the second Met (amino acid 296) of G7a (results not shown). This means that there is no identity in yeast valyl-tRNA synthetase with the 154 amino acid *N*-terminal domain of G7a which contains two stretches of

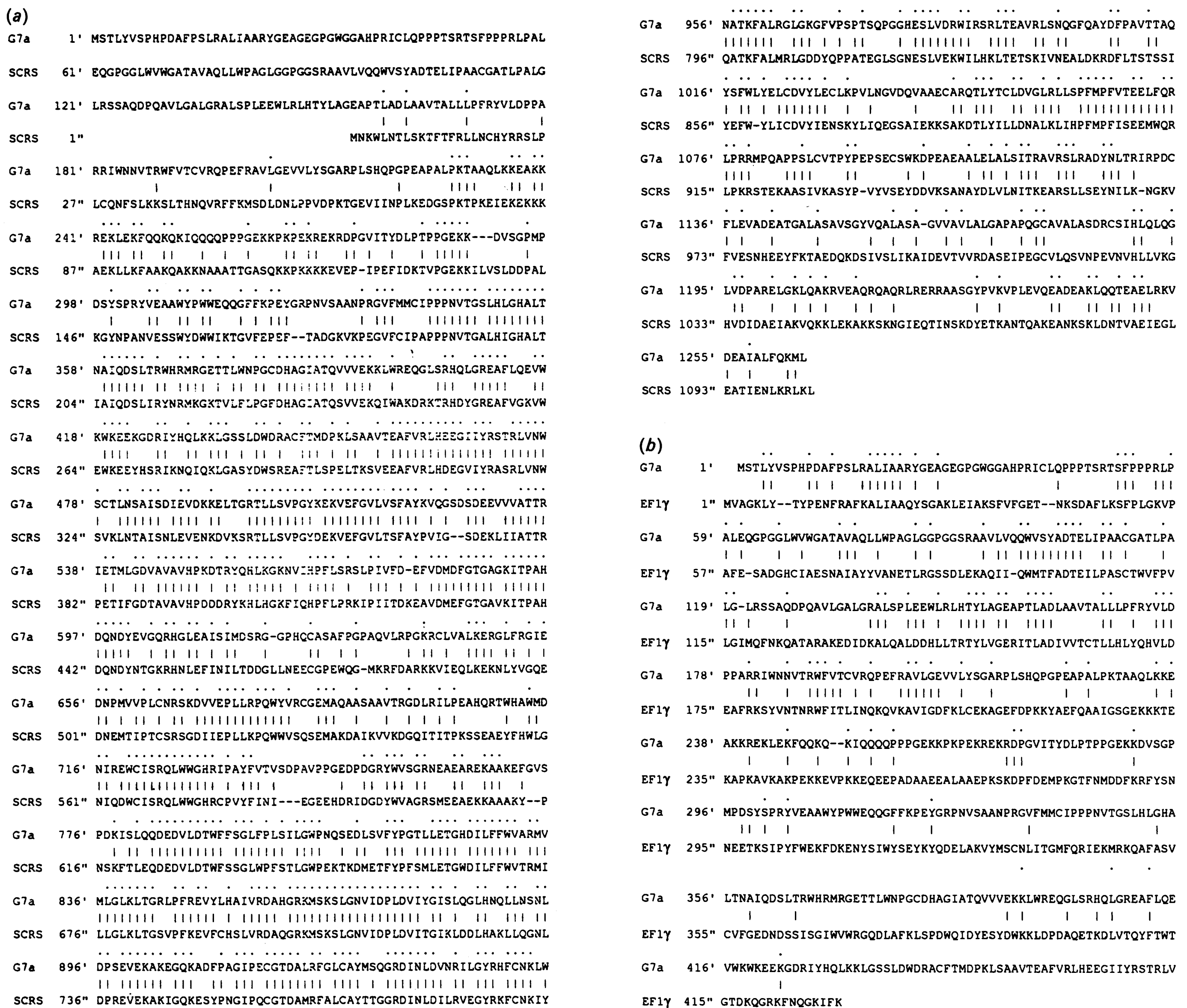

Fig. 5. Alignment of the G7a protein with *S. cerevisiae* valyl-tRNA synthetase (*a*) and brine shrimp elongation factor 1 γ-chain (*b*)

The deduced amino acid sequence of G7a is aligned with the primary sequence of the *S. cerevisiae* valyl-tRNA synthetase (SCRS) and brine shrimp elongation factor 1 γ-chain (EF1γ). The vertical line (|) represents chemically equivalent amino acids. Identical amino acids are marked by a dot above the sequence.

hydrophobic amino acids (65–85 and 99–119). However, the basic-amino-acid-rich segment (amino acids 228–278) is present in both G7a and the *S. cerevisiae* enzyme, but is not present in the *E. coli* enzyme. Besides, the unique 154-amino-acid *N*-terminal domain in G7a shows strong sequence similarity with brine shrimp elongation factor 1 γ-chain, especially in the most hydrophobic region (amino acids 99–121). It is likely that this segment of the protein is responsible for the hydrophobic interaction between elongation factor 1 β- and γ-chains, and causes high-molecular-mass aggregates to form between mammalian valyl-tRNA synthetase and elongation factor 1H. As the molecular mass of valyl-tRNA synthetase increases from prokaryotes to mammals, and from the alignment of G7a and *E. coli* valyl-tRNA synthetase, we propose that the *N*-terminal domain (amino acids 1–295) which is absent in *E. coli* valyl-tRNA synthetase, has been acquired recently during evolution. This is in accordance with eukaryotic valyl-tRNA synthetase having more functions than that of prokaryotes.

The function of the basic *N*-terminal domains found in both human and yeast valyl-tRNA synthetases, but not in that of prokaryotes, is still unclear. There are two different forms of arginyl-tRNA synthetase in rat liver, one existing as a monomer and the other as a higher-molecular-mass complex (Vellekamp & Deutscher, 1987). The difference between these two forms is due to a basic *N*-terminal extension in the higher-molecular-mass complex form of arginyl-tRNA synthetase which is required for complex formation. The basic domain in human valyl-tRNA synthetase cannot be solely responsible for the complex formation with elongation factor 1H, since the yeast valyl-tRNA synthetase, which also has this domain, is isolated as a monomer. Comparison of the amino acid sequences of human and yeast aspartyl-tRNA synthetases (Jacobo-Monina *et al.*, 1989) has revealed

```
BLOCK 1
G7a   340  CIPPPNVTGSLHLGHALTNAI 360
SCVS  186  PA.......A..I.....I.. 206
ECVS   38  M........M......FQQT.  58
BSVS   45  V........K......WDTTL  65
SCLS   27  LCQF.YPS.A..I..LRVYV.  47

BLOCK 2
G7a   379  PGCDHAGIATQVVVEKKLW  397
SCVS  225  ..F........S....QI.  243
ECVS   77  V.T........M...R.IA   95
BSVS   86  ..M........AK..E..R  102
SCLS   66  M.W.AF.LPAENAAIERSI   84

BLOCK 3
G7a   527  SDEEVVVATTRIETMLGDVAVAVHPKDTRYQHLKGK  562
SCVS  371  ...KLII....P..IF..T......D.D..K..H..  406
ECVS  213  GKDYL......P..L...TG...N.E.P..KD.I..  248
BSVS  213  GSGFIE.....P......T......D.E..K..I..  248
SCLS  230  KF.NLI.F...P..LFAVQY..LALDHPIV.KYCEE  265

BLOCK 4
G7a   581  VDMDFGTGAGKITPAHDQNDYEVGQRHGLEAI  612
SCVS  426  ...E.....V...........NT.K..N..F.  457
ECVS  267  A..EK...CV.......F......K..A.PM.  298
BSVS  267  ...E..S..V.......P..F.I.N..N.PR.  298
SCLS  319  SSYGSAPS.VMGC.G..NR.F.FWQTNCPGEH  350

BLOCK 5
G7a   716  NIREWCISRQLWWGHRIPAYF  736
SCVS  561  ..QD...........C.V..  581
ECVS  419  D.QD...............WY  439
BSVS  402  ...D...............W.  422
SCLS  422  KI.D.L....RY..TP..IIH  442

BLOCK 6
G7a   846  PFREVYLHAIVRDAHGRKMSKSLGNVIDPLDVIYGISL  883
SCVS  686  ..K..FC.SL....Q..................T..K.  723
ECVS  537  ..HT..MTGLI..DE.Q.....K........MVD....  574
BSVS  508  .EKD..I.GL....Q..........GV..M...DQYGA  545
SCLS  612  TTVIIKSNGK.PVVSYE.....KY.GA..NEC.LRHGP  639
```

Fig. 6. Alignment of the conserved sequence motifs between human valyl-tRNA synthetase and other tRNA synthetases

Alignment of human valyl-tRNA synthetase (G7a) with *S. cerevisiae* valyl-tRNA synthetase (SCVS), *E. coli* valyl-tRNA synthetase (ECVS), *B. stearothermophilus* valyl-tRNA synthetase (BSVS) and *S. cerevisiae* leucyl-tRNA synthetase (SCLS). The six motifs are shared by all of the tRNA synthetases which charge the aliphatic amino acids. The His-Ile-Gly-His (HIGH) and Lys-Met-Ser-Lys-Ser (KMSKS) regions are highlighted in bold. A dot represents an identical amino acid. The number on either side represents the position of the *N*-terminal (left) and *C*-terminal (right) amino acid residues in the protein sequence.

that the basic positively charged helices in the yeast synthetase have evolved to neutral amphiphilic helices in this mammalian synthetase, which may account for the association of the human synthetase in the complex form. From these findings, it appears that the domain responsible for the higher-molecular-mass complex formation in different tRNA synthetases is variable.

Increasing evidence suggests that some tRNA synthetases are involved in mRNA splicing (Akins & Lambowitz, 1987; Herbert *et al.*, 1988; Cherniack *et al.*, 1990). Since *S. cerevisiae* leucyl-tRNA synthetase, a class I enzyme charging an aliphatic amino acid to tRNA, also has this ability, pairwise comparison was carried out with valyl-tRNA synthetase (Fig. 6). The six conserved blocks could be found in all the tRNA synthetases which charge aliphatic amino acids (Herbert *et al.*, 1988), but the typical 'zinc-finger' motif which appears in block 4 of *S. cerevisiae* mitochondrial leucyl-tRNA synthetase is absent from all of the valyl-tRNA synthetases, since Cys (at positions 331 and 346) and His (at position 350) are replaced by amino acids that cannot act as metal ligands. Thus there is no equivalent motif in valyl-tRNA synthetase that has the potential for intron-splicing activity.

Whereas bacterial organisms typically have just one aminoacyl-tRNA synthetase for each amino acid, in yeast some tRNA synthetases have two different forms, one in the cytosol and the other in mitochondria. Among these is histidyl-tRNA synthetase (HTS1), which is encoded by a single-copy gene but has two in-frame translation start sites located 60 bp apart. Translation from these two start sites results in different forms of histidyl-tRNA synthetase (Natsoulis *et al.*, 1986). One set of HTS1 transcripts (long) initiates upstream of both ATG codons, and the other set (short) initiates between the two ATG codons, and therefore contains only the downstream ATG. The longer mRNA encodes the mitochondrial synthetase, which has 20 more *N*-terminal amino acids, and the shorter mRNA encodes the cytoplasmic histidyl-tRNA synthetase. The nuclear and mitochondrial forms of the tryptophanyl- and threoninyl-tRNA synthetases, on the other hand, are encoded by different genes (Myers *et al.*, 1985; Pape & Tzagoloff, 1985; Pape *et al.*, 1985). However, based on Southern blotting data and the human valyl-tRNA synthetase amino acid sequence, neither of the above features are found in this gene, so there is probably only one form of this synthetase in mammalian cells.

The valyl-tRNA synthetase gene is the first tRNA-synthetase gene whose position in the human genome has been precisely mapped, though the human threonyl-tRNA synthetase gene has been mapped to the short arm of human chromosome 5 (Kontis & Arfin, 1989). The location of the valyl-tRNA synthetase gene in the class III region of the MHC emphasizes that this segment of DNA, and possibly also the class I and class II regions (Trowsdale *et al.*, 1990; Spies *et al.*, 1990), will contain a heterogeneous collection of genes, a significant number of which are not directly involved in the immune response.

We thank Dr. Ken Reid for critical review of the manuscript, Mark Olavesen for help in isolating the cDNA clones, Sandra Smith for excellent technical assistance, and Ken Johnson for photographic help. S.-L. H. holds a grant from the Ministry of Education, Taiwan, Republic of China.

REFERENCES

Akins, R. A. & Lambowitz, A. M. (1987) Cell **50**, 331–345
Bec, G. & Waller, J. (1989) J. Biol. Chem. **264**, 21138–21143
Bec, G., Kerjan, P., Zha, X. D. & Waller, J. (1989) J. Biol. Chem. **264**, 21131–21137
Bird, A. P. (1986) Nature (London) **321**, 209–213
Bird, A. P. (1987) Trends Genet. **3**, 342–347
Carroll, M. C., Campbell, R. D., Bentley, D. R. & Porter, R. R. (1984) Nature (London) **307**, 237–241
Carroll, M. C., Campbell, R. D. & Porter, R. R. (1985) Proc. Natl. Acad. Sci. U.S.A. **82**, 521–525
Carroll, M. C., Katzman, P., Alicot, E. M., Koller, B. H., Geraghty, D. E., Orr, H. T., Strominger, J. L. & Spies, T. (1987) Proc. Natl. Acad. Sci. U.S.A. **84**, 8535–8539
Cherniack, A. D., Garriga, G., Kittle, J. D., Akins, R. A. & Lambowitz, A. M. (1990) Cell **62**, 745–755
Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. (1979) Biochemistry **18**, 5294–5299
Davis, M. M. & Bjorkman, P. J. (1988) Nature (London) **334**, 395–402
Dunham, I., Sargent, C. A., Trowsdale, J. & Campbell, R. D. (1987) Proc. Natl. Acad. Sci. U.S.A. **84**, 7237–7241
Dunham, I., Sargent, C. A., Kendall, E. & Campbell, R. D. (1990) Immunogenetics **32**, 175–182
Eisenberg, D., Schwarz, E., Komaromy, M. & Wall, R. (1984) J. Mol. Biol. **179**, 125–142
Eriani, G., Delarue, M., Poch, O., Gangloff, J. & Moras, D. (1990) Nature (London) **347**, 203–206
Feinberg, A. P. & Vogelstein, B. (1984) Anal. Biochem. **137**, 266–267
Fourney, R. M., Miyodoshi, J., Day, R. F. & Paterson, M. C. (1988) Focus **10**, 5–7

Gardiner-Garden, M. & Frommer, M. (1987) J. Mol. Biol. **196**, 261–282
Heck, J. D. & Hatfield, G. W. (1988) J. Biol. Chem. **263**, 868–877
Herbert, C. J., Labouesse, M., Dujardin, G. & Slonimski, P. P. (1988) EMBO J. **7**, 473–483
Jacobo-Monina, A., Peterson, R. & Yang, D. C. H. (1989) J. Biol. Chem. **264**, 16608–16612
Jordana, X., Chatton, B., Paz-Weisshaar, M., Buhler, J. M., Cramer, F., Ebel, J. P. & Fasiolo, F. (1987) J. Biol. Chem. **262**, 7189–7194
Kellermann, D., Tonetti, H., Brevet, A., Mirande, M. & Pailliez, J. P. (1982) J. Biol. Chem. **257**, 11041–11048
Kendall, E., Sargent, C. A. & Campbell, R. D. (1990) Nucleic Acids Res. **18**, 7251–7257
Kontis, K. J. & Arfin, S. M. (1989) Mol. Cell. Biol. **9**, 1832–1838
Kozak, M. (1984) Nucleic Acids Res. **11**, 857–872
Maniatis, T., Fritsch, E. F. & Sambrook, L. (1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
Melton, D., Krieg, P. A., Rebagliati, M. R., Maniatis, T., Zinn, K. & Green, M. R. (1984) Nucleic Acids Res. **12**, 7035–7056
Milner, C. M. & Campbell, R. D. (1990) Immunogenetics **32**, 175–182
Myers, A. M., Pape, L. K. & Tzagoloff, A. (1985) EMBO J. **4**, 2087–2092
Natsoulis, G., Hilger, F. & Fink, G. R. (1986) Cell **46**, 235–274
Pape, L. K. & Tzagoloff, A. (1985) Nucleic Acids Res. **13**, 6171–6183
Pape, L. K., Koerner, T. J. & Tzagoloff, A. (1985) J. Biol. Chem. **260**, 15632–15370
Ragoussis, J., Monaco, A., Mockridge, I., Kendall, E., Campbell, R. D. & Trowsdale, J. (1991) Proc. Natl. Acad. Sci. U.S.A. **88**, 3753–3757
Sanger, F., Nicklen, S. & Coulson, A. R. (1979) Proc. Natl. Acad. Sci. U.S.A. **74**, 5463–5467
Sargent, C. A., Dunham, I. & Campbell, R. D. (1989) EMBO J. **8**, 2305–2312
Schimmel, P. R. (1987) Annu. Rev. Biochem. **56**, 125–158
Schimmel, P. R. (1989) Adv. Enzymol. **63**, 233–271
Schimmel, P. R. & Soill, D. (1979) Annu. Rev. Biochem. **48**, 601–648
Seed, B. (1987) Nature (London) **329**, 840–842
Simmons, D. & Seed, B. (1988) J. Immunol. **141**, 2797–2800
Southern, E. M. (1975) J. Mol. Biol. **98**, 503–517
Spies, T., Morton, C. C., Nedospasov, S. A., Fiers, W., Pious, D. & Strominger, J. L. (1986) Proc. Natl. Acad. Sci. U.S.A. **83**, 8699–8702
Spies, T., Bresnahan, M. & Strominger, J. L. (1989) Proc. Natl. Acad. Sci. U.S.A. **86**, 8955–8958
Spies, T., Bresnahan, M., Bahram, S., Arnold, D., Blank, G., Mellins, E., Pious, D. & DeMars, R. (1990) Nature (London) **348**, 744–747
Staden, R. (1986) Nucleic Acids Res. **14**, 217–231
Strachan, T. (1987) Br. Med. Bull. **43**, 214–217
Trowsdale, J. (1987) Br. Med. Bull. **43**, 15–36
Trowsdale, J., Hanson, I., Mockridge, I., Beck, S., Townsend, A. & Kelley, A. (1990) Nature (London) **348**, 741–744
Tsui, F. W. L. & Siminovitch, L. (1987) Nucleic Acids Res. **15**, 3349–3367
Ussery, M. A., Tanaka, W. K. & Hardesty, B. (1977) Eur. J. Biochem. **72**, 491–500
Vellekamp, G. & Deutscher, M. P. (1987) J. Biol. Chem. **262**, 9927–9930
White, P. C., Grossberger, D., Onufer, B., Chaplin, D. D., New, M. I., Dupont, B. & Strominger, J. L. (1985) Proc. Natl. Acad. Sci. U.S.A. **82**, 1089–1093

Received 27 February 1991/19 April 1991; accepted 22 April 1991