The Human Lysyl-tRNA Synthetase Gene Encodes Both the Cytoplasmic and Mitochondrial Enzymes by Means of an Unusual Alternative Splicing of the Primary Transcript*

Received for publication, July 14, 2000, and in revised form, August 16, 2000
Published, JBC Papers in Press, August 21, 2000, DOI 10.1074/jbc.M006265200

Elena Tolkunova, Hyejeong Park, Jun Xia, Michael P. King†, and Edgar Davidson

From the Department of Biochemistry and Molecular Pharmacology, Thomas Jefferson University, Philadelphia, Pennsylvania 19107

Two cDNAs encoding human lysyl-tRNA synthetase have been identified. One encodes the cytoplasmic form of the enzyme identified previously. The second cDNA contains the same sequence but with a 180-bp insertion at the 5′-end of the mRNA. This results in a predicted protein whose carboxyl 576 amino acids are identical to those of the cytoplasmic enzyme but with a different amino terminus of 99 amino acids that contains a putative mitochondrial targeting sequence. Expression of the two lysyl-tRNA synthetase-green fluorescent protein gene fusions in a human cell line confirmed that the cytoplasmic form was targeted to the cytoplasm and the mitochondrial form to mitochondria. The genomic lysyl-tRNA synthetase gene consisted of 15 exons. The two isoforms were created by alternative splicing of the first three exons of the gene. The cytoplasmic form was created by splicing exon 1 to exon 3. The inclusion of exon 2 between exons 1 and 3 produced an mRNA encoding the mitochondrial isoform with an additional upstream small open reading frame, consisting mainly of a portion of the 5′ coding region of the cytoplasmic isoform. This is the first example of mitochondrial targeting sequence being encoded on the second exon of a gene. Ribonuclease protection analysis showed that the mRNA encoding the cytoplasmic isoform makes up approximately 70%, and the mitochondrial isoform approximately 30%, of the mature transcripts from the lysyl-tRNA synthetase gene. The mitochondrial form of the enzyme, purified after expression in Escherichia coli, aminoacylated in vitro transcripts corresponding to both the cytoplasmic and mitochondrial tRNA<sup>3-5′</sup>, despite the difference in the discriminator base sequence in the acceptor stems of these tRNAs.

The tRNA synthetases are a group of enzymes responsible for aminoacylating tRNAs with appropriate amino acids. In recent years, a considerable amount of information has been obtained about their functions, but there are still a number of uncharacterized members of this diverse group of enzymes.

Genes encoding lysyl-tRNA synthetases have been cloned from a number of organisms, and crystal structures have been obtained for one of the two Escherichia coli lysyl-tRNA synthetases, LysU (1), and the Thermus thermophilus lysyl-tRNA synthetase (2). The majority of the lysyl-tRNA synthetases are class II enzymes; however, some archaeabacteria have a class I lysyl-tRNA synthetase (3). The cDNA encoding the human cytoplasmic lysyl-tRNA synthetase has been described previously, and the protein purified after expression in E. coli has been characterized (4). In mammals, the cytoplasmic form is a component of a synthetase complex that contains an additional seven tRNA synthetase activities (5, 6). The amino-terminal region of the protein has been implicated in the interactions with the multienzyme complex (5, 6). The gene encoding human lysyl-tRNA synthetase has been localized to chromosome 16q23-q24 (7).

Since protein translation occurs in both the cytosol and mitochondria of eukaryotic cells, the cell generally requires tRNA synthetase activities for each amino acid in both subcellular locations. In some cases, two distinct genes encode the cytoplasmic and mitochondrial isoforms, e.g. human histidyl-tRNA synthetase (8). Alternatively, one gene may encode both forms of the protein. In the case of glycyl-tRNA synthetase, this is achieved by using alternative transcription initiation sites, with the furthest upstream transcript start site resulting in the addition of a mitochondrial targeting sequence to the amino terminus of the protein (9, 10). A number of the human mitochondrial tRNA synthetases have yet to be identified and characterized. In yeast, the cytoplasmic and mitochondrial isoforms of the lysyl-tRNA synthetase are encoded by two distinct genes (11, 12). To identify and clone the gene encoding the mitochondrial form of human lysyl-tRNA synthetase, we searched the NCBI human EST<sup>1</sup> data base to identify potential ESTs that were homologous, but not identical, to the human lysyl-tRNA synthetase. We identified a human EST cDNA sequence that appeared to be an altered form, or alternative splicing product, of the cytoplasmic form of the gene (AA356156). We investigated the possibility that the alternative form of the gene encoded the mitochondrial lysyl-tRNA synthetase.

**EXPERIMENTAL PROCEDURES**

Overexpression and Purification of Lysyl-tRNA Synthetase—Plasmid pM368 (4), containing the full-length cytoplasmic lysyl-tRNA synthetase cloned into a derivative of E. coli expression vector pET19b, was kindly given to us by Dr. Kiyotaka Shiba (Japanese Foundation for Cancer Research, Tokyo, Japan). The cDNA encoding the mitochondrial lysyl-tRNA synthetase was cloned into E. coli expression vector pET24d (Novagen). The second codon in the cDNA, encoding leucine, was changed to one encoding alanine to improve protein stability following expression in E. coli, in accordance with the N-ends rule (13). These constructs were expressed in E. coli BL21(DE3) Codon Plus (RIL).

<sup>1</sup>The abbreviations used are: EST, expressed sequence tag; PCR, polymerase chain reaction; GFP, green fluorescent protein; bp, base pair(s); nt, nucleotide(s); aa, amino acid(s); uORF, upstream open reading frame.
In Vitro Aminoacylation Studies—Cytoplasmic tRNA\(^{\text{Lys}}\) was synthesized by \(\text{in vitro}\) run off transcription (15), using pLysF119 (14) as the template. Since the T7 RNA polymerase initiates with G, and since the mitochondrial tRNA\(^{\text{Lys}}\) begins with C, the mitochondrial tRNA\(^{\text{Lys}}\) gene was cloned so that tRNA\(^{\text{Lys}}\) was preceded by a hammerhead ribozyme (16). This plasmid was used as a template to make a run-off transcript that was then incubated to release a transcript starting with the correct nucleotide. The resulting tRNAs were purified by electrophoresis through polyacrylamide-urea gels. Aminoacylation was performed at 37 °C as described previously (15) using 1 \(\mu\)Ci/mmol [\(^{3}\text{H}\)]lysine (89 Ci/mmol; PerkinElmer Life Sciences), and 0.9–88 nm mitochondrial lysyl-tRNA synthetase in a total reaction volume of 20 \(\mu\)L. At 3-min intervals, 4–\(\mu\)L aliquots were spotted on 3MM filter paper and soaked in cold 5% trichloroacetic acid to terminate the reaction. The amount of chargeable tRNA in the purified transcript was calculated following a time course. The reaction was carried out using 1.8 \(\mu\)M mitochondrial lysyl-tRNA synthetase with 0.2 \(\mu\)M cytoplasmic tRNA\(^{\text{Lys}}\)\(^{\text{59}}\) transcript or 0.7 \(\mu\)M mitochondrial tRNA\(^{\text{Lys}}\)\(^{\text{59}}\) transcript and using cytoplasmic lysyl-tRNA synthetase with 0.4 \(\mu\)M cytoplasmic tRNA\(^{\text{Lys}}\)\(^{\text{59}}\) transcript or 0.8 \(\mu\)M mitochondrial tRNA\(^{\text{Lys}}\)\(^{\text{59}}\) transcript.

Quantitation of the Relative Amounts of mRNA Transcripts Encoding the Cytoplasmic or Mitochondrial Forms of Lysyl-tRNA Synthetase—The relative amounts of each transcript in total cellular RNA extracts from 143B cells were quantitated using the Ribonuclease Protection Assay (RPA) system RPA II (Ambion) according to the manufacturer’s suggested protocol. A gene fragment corresponding to exon 2 (121 bp, encoding the mitochondrial targeting sequence) and part of exon 3 (78 bp, encoding a region common to both cytoplasmic and mitochondrial proteins) was subcloned into pGEM-Teasy. The resulting plasmid was used as a template for \(\text{in vitro}\) transcription in the presence of \([\alpha-\text{\textsuperscript{32}P}]\)UTP to give a labeled fragment of 363 nt. Different amounts of total RNA were tested in incubations with a large excess of the 363-nt transcript. After nuclease digestion, the protected fragments were separated by electrophoresis through a 6% polyacrylamide gel. Signals corresponding to the mitochondrial and cytoplasmic transcripts were quantitated by a PhosphorImager (Molecular Dynamics, Inc., Sunnyvale, CA).

Comparison of Lysyl-tRNA Synthetase Genes among Eukaryotic Species—We used the following GenBank™ EST and genomic sequences to compare the genomic or mRNA 5‘ sequences encoding lysyl-tRNA synthetase: *Caenorhabditis elegans* mitochondrial isoform mRNA C09136.1, cytoplasmic isoform mRNA AV188647.1, genomic sequence U41105.1; *Drosophila melanogaster* mitochondrial isoform mRNA AE516192.1, cytoplasmic isoform mRNA AE003447.1; mitochondrial isoform mRNA AW258396, cytoplasmic isoform mRNA W53766; *zebrafish* (Danio rerio) mitochondrial isoform mRNA AW421393, and cytoplasmic isoform mRNA, AW421565.1.
cDNAs encoding both the cytoplasmic and mitochondrial isoforms of mitochondrial lysyl-tRNA synthetase were amplified by reverse transcriptase-PCR. The sequences were identical except for a 180-bp insertion close to the 5'-end to produce the putative mitochondrial isoform (see Fig. 1A). This insertion introduces a sequence encoding Lys-Trp-Trp-STOP into the cytoplasmic isoform reading frame after the codon for Lys^{17}. The next ATG is 24 bp downstream from the stop codon and presumably is the initiation codon of the mitochondrial isoform. The reading frame of the mitochondrial isoform encodes 48 amino acids before resuming the coding sequence of the cytoplasmic isoform at amino acid Glu^{22}. The 48 amino acids have the characteristics of a mitochondrial targeting sequence. The program Mitoprot (17) predicts a mitochondrial location for this protein. The amino acids contained in both mitochondrial and cytoplasmic lysyl-tRNA synthetase are summarized under "Experimental Procedures." BAC clones were screened by Southern analysis, using a probe corresponding to the region of the cDNA encoding the putative mitochondrial targeting sequence. Eight clones were identified that contained the 5’-end of the lysyl-tRNA synthetase gene on a 13-kb BamHI fragment. One, BAC clone 2256 J17, was selected for more detailed analyses. A combination of Southern and DNA sequence analysis confirmed that the 13-kb BamHI fragment contained the first four exons of the lysyl-tRNA synthetase gene. A partial genomic sequence of the lysyl-tRNA synthetase locus was determined by direct sequencing of the BAC clone 2256 J17 and the genomic library clone. The introns 1, 2, 3, 4, 7, and 8 were not fully sequenced, but their sizes were estimated by restriction mapping, Southern analysis, and PCR amplification of the BAC clones and genomic DNA.

The complete gene contained 15 exons and extended over approximately 20 kb of the genome (Fig. 2). The intron-exon borders are shown in Table I. A comparison of the genomic sequence with those of the two forms of cDNA obtained showed that the two forms of lysyl-tRNA synthetase were obtained by alternative splicing of the first three exons of the gene (Fig. 3). The first exon encodes the amino-terminal 20 amino acids of the cytoplasmic form of the enzyme. The second exon encodes an open reading frame with an initiating methionine and 48 amino acids including a region with the characteristics of a mitochondrial targeting sequence. The cytoplasmic form of the enzyme is encoded by an mRNA containing the first exon spliced to the third exon, creating the cDNA for the cytoplasmic form of the enzyme described previously (4). The mitochondrial form is encoded by an mRNA whose 5’ region consists of the first three exons spliced consecutively. The reading frame created by the first exon is terminated by a stop codon in the
The Human Lysyl-tRNA Synthetase Gene

second exon. The initiating methionine of the mitochondrial isoform is 24 bp downstream from this stop codon (Fig. 1).

Upstream from the initiation ATG of the human cytoplasmic lysyl-tRNA synthetase gene, on the opposite strand, is a reading frame of 1200 bp starting at position −243. Investigation of the corresponding cDNA in the NCBI EST database allowed us to construct a putative full-length cDNA. We used this information to design primers that were used to clone the mRNA by PCR from cDNA.

The protein encoded by the upstream reading frame, which we designated KRU (lysyl-tRNA synthetase reading frame upstream) in Fig. 2, had no homologies to known proteins. A ProfileScan analysis for protein motifs identified a BRCT domain (residues 78–101) and a Myb domain (residues 128–188). This protein has since been identified as a human ortholog of Rap1 (Ref. 18; GenBank™ accession number AF262988), a protein, localized at telomeres, that affects telomere length.

The region of chromosome 16 containing the lysyl-tRNA synthetase and Rap1 is represented by “working draft” sequences in GenBank™ (accession numbers AC025287 and AC011934). There is a region of 243 bp between the two initiation codons for lysyl-tRNA synthetase and KRU/Rap1. This region lacks a conventional TATA sequence, as is characteristic of housekeeping genes, but presumably contains a bidirectional promoter and is predicted to contain several Sp1 binding domains in both orientations. A ribonuclease protection assay analysis of mRNA isolated from human 143B cells showed that the major initiation site for the lysyl-tRNA synthetase transcript occurs at the position corresponding to −31 (data not shown).

Analysis of Other Eukaryotic Lysyl-tRNA Synthetases—Analysis of the GenBank™ EST and genomic sequence data bases (see “Experimental Procedures” for the appropriate accession numbers) showed that it is likely that one gene provides both cytoplasmic and mitochondrial isoforms of lysyl-tRNA synthetase in mouse, C. elegans, D. melanogaster, and zebrafish (D. rerio) (Fig. 1B). In each case, the cytoplasmic form would be created by the exclusion of the second exon, containing a mitochondrial targeting sequence during the splicing of the primary transcript.

Both the C. elegans and D. melanogaster genomic lysyl-tRNA synthetase sequences are predicted to contain four exons. In C. elegans, two of the three positions of the introns in the coding sequence are absolutely conserved with an equivalent human intron; the third is within 3 aa (10 bp) of the equivalent site in human DNA. Exon 1 contains the coding regions of the amino termini of both the cytoplasmic and mitochondrial forms, since the 5′ cytoplasmic/mitochondrial boundary region of the mRNA is not interrupted by an intron in the genome. To form the mRNA encoding the cytoplasmic form, splicing proceeds from a point midway in exon 1 to exon 2. In Drosophila, the locations of the introns are conserved with the human gene only for introns 1 and 2. The sequences in the EST data base suggest that, in Drosophila, the incorporation of the mitochondrial amino terminus might be accomplished by initiation of transcription at exon 2.

Subcellular Localization of the Isoforms of Human Lysyl-tRNA Synthetase—To confirm the identities of the mRNA species predicted to encode the two isoforms of lysyl-tRNA synthetase, they were cloned into mammalian expression vector pEFmyc/cyto/GFP so that a gene encoding GFP was added, in frame, to the 3′-ends of both mRNAs. These constructs were then transfected into the human osteosarcoma cell line 143B for transient expression studies. Cells were fixed 24 h after transfection and examined for GFP fluorescence or immunostained. As expected, the cytoplasmic lysyl-tRNA synthetase-GFP fusion construct produced a diffuse, cellwide fluorescence pattern (Fig. 4A). The putative mitochondrial lysyl-tRNA synthetase-GFP fusion construct resulted in a punctate pattern of fluorescence characteristic of a mitochondrial distribution (Fig. 4C). To confirm that this localization was mitochondrial, cells expressing the lysyl-tRNA synthetase-GFP fusion protein were immunostained with an antibody directed against the inner mitochondrial membrane protein COX I and a rhodamine-labeled secondary antibody (Fig. 4B). The GFP and rhodamine staining co-localized. Thus, the two lysyl-tRNA synthetase mRNAs encoded enzymes that were destined for two distinct subcellular locations, the mitochondria and the cytosol.

Table I

| Intron No. | Estimate BP | HTGS | 5′ Phase | 3′ Phase | Exon No. | Size BP |
|---|---|---|---|---|---|---|
| 1 | ~3000 | 3115 | AAG GTA GTAGT | gAG GTAGT GGTG | 1 | 9 |
| 2 | ~2500 | 2559 | gAG GTAGT GGTG | cAG GTAGG CCA | 2 | 189 |
| 3 | ~1350 | 1213 | aATG GTAGGT | cAG GTAGG CCA | 3 | 160 |
| 4 | ~3400 | 355 | gCA GTG Aaa | cAG GTAGG CCA | 1 | 166 |
| 5 | 355 | 355 | cAG GTAGG CCA | tAG GTAGG CCA | 2 | 167 |
| 6 | 106 | 106 | cAG GTAGG CCA | tAG GTAGG CCA | 0 | 126 |
| 7 | ~1800 | 1386 | aAG GTAGA Gaa | tAG GTAGG CCA | 0 | 120 |
| 8 | ~2160 | 2317 | tAG GTAGG CCA | tAG GTAGG CCA | 0 | 163 |
| 9 | 103 | 103 | tAG GTAGG CCA | cAG GAAGT GGTG | 1 | 10 |
| 10 | 162 | 162 | cAG GAAGT GGTG | tAG GTAGG CCA | 1 | 115 |
| 11 | 659 | 659 | cAG GTAGG CCA | tAG GTAGG CCA | 0 | 85 |
| 12 | ~900 | 881 | aATG GAAGT GGTG | tAG GTAGG CCA | 2 | 134 |
| 13 | 697 | 702 | aAG GTAGG CCA | cAG GAAGT GGTG | 0 | 138 |
| 14 | 576 | 575 | cAG GAAGT GGTG | aAG GAAGT GGTG | 0 | 260 |
plasmic isoform. Exon 2, encoding the amino terminus of the mitochondrial lysyl-tRNA synthetase, is represented by a white box. Striped boxes represent exons 3–15. Inclusion of exon 2 in a transcript results in the major open reading frame encoding a lysyl-tRNA synthetase with a mitochondrial targeting sequence at the amino terminus of the protein, encoded on exon 2. Exclusion of exon 2 results in a transcript encoding the cytoplasmic isoform of the protein, with the amino terminus encoded on exon 1.

Fig. 3. Production of two isoforms of the human lysyl-tRNA synthetase by alternative splicing of the first three exons of the gene. Shown is a diagram that illustrates the alternative splicing that produces the two forms of human lysyl-tRNA synthetase transcripts. A white box represents exon 1, encoding the amino terminus of the cytoplasmic isoform. Exon 2, encoding the amino terminus of the mitochondrial isoform, is represented by a black box. Striped boxes represent exons 3–15. Inclusion of exon 2 in a transcript results in the major open reading frame encoding a lysyl-tRNA synthetase with a mitochondrial targeting sequence at the amino terminus of the protein, encoded on exon 2. Exclusion of exon 2 results in a transcript encoding the cytoplasmic isoform of the protein, with the amino terminus encoded on exon 1.

Fig. 4. Subcellular localization of human cytoplasmic and mitochondrial lysyl-tRNA synthetase-GFP fusion proteins. Human cell line 143B was transiently transfected with constructs containing the two full-length lysyl-tRNA synthetase coding regions fused with GFP. A, human cell line 143B was transiently transfected with a pEF/myc/cyto/GFP construct containing the full-length human cytoplasmic lysyl-tRNA synthetase coding region fused with GFP. Diffuse GFP fluorescence was observed throughout the cytoplasm. Note the exclusion of the majority of the fluorescence from the nucleus. B and C, 143B cells were transfected with a pEF/myc/cyto/GFP construct containing the full-length human mitochondrial lysyl-tRNA synthetase coding region fused with GFP. Shown in B is the immunofluorescence pattern of staining using a primary antibody directed against the mitochondrial COX I protein and a rhodamine-labeled secondary antibody. The immunofluorescence exhibits a punctate distribution in the cytoplasm characteristic of a location in mitochondria. The direct fluorescence of GFP in the same field of view is shown in C. In cells expressing the mitochondrial lysyl-tRNA synthetase-GFP construct, the GFP fluorescence co-localizes with the immunofluorescence of COX I. This indicates that the mitochondrial lysyl-tRNA synthetase is localized in mitochondria.

Quantitation of the Relative Amounts of the Cytoplasmic and Mitochondrial Lysyl-tRNA Synthetase Transcripts—The relative amounts of the two lysyl-tRNA synthetase transcripts were quantitated by a ribonuclease protection assay using radiolabeled sense and antisense RNA probes designed using a cDNA sequence corresponding to a portion of the mitochondrial lysyl-tRNA synthetase isoform (Fig. 5A). The mitochondrial form of lysyl-tRNA synthetase was predicted to protect a fragment of 199 nt, while the cytoplasmic form was predicted to protect a fragment of 78 nt. The 363-nt antisense probe (Fig.

Fig. 5. Quantitation of transcripts encoding isoforms of human lysyl-tRNA synthetase. A, a schematic diagram indicating the location of the 363-nt antisense RNA probe used for the ribonuclease protection assay is shown. The region of the cDNA representing exon 1 is white, and the region of the cDNA representing exon 2 is black. The remainder of the cDNA is striped. The probe contained sequences unrelated to lysyl-tRNA synthetase at both its 5′- and 3′-ends as indicated by the stippled boxes. B, after RNase treatment the protected RNA probes were separated by electrophoresis through a 6% polyacrylamide gel and quantitated by a PhosphorImager. The undigested probe is shown in lane 1. 10% of the amount of probe used in the assay was loaded in this lane; the exposure shown for lane 1 is shorter than that for lanes 2 and 3. Lanes 2 and 3 show the result of hybridizing the probe with total RNA isolated from osteosarcoma cell line 143B (lanes 2, 32 μg; lane 3, 16 μg) followed by RNase digestion. The regions protected by the mitochondrial transcript (mtKRS, 199 nt) and by the cytoplasmic transcript (cytKRS, 78 nt) are indicated. Sizes were calculated by comparison with a 32P-end-labeled 100-bp ladder (data not shown).

5B, lane 1) was hybridized with total RNA isolated from human osteosarcoma cell line 143B, followed by digestion with RNases A and T1. (Fig. 5B, lanes 2 and 3). The two fragments obtained after RNase digestion were estimated to be 202 and 67 nt, consistent with the predicted lengths. The sense probe was not protected by incubation with total RNA (data not shown). The ratio of the two transcripts remained constant over a 10-fold range of the amount of total RNA used (not shown). After correction for the number of radiolabeled UTPs incorporated into each probe, it was calculated that the mRNA encoding the cytoplasmic lysyl-tRNA synthetase accounted for approximately 69 ± 3% of the total lysyl-tRNA synthetase mRNA, while the transcript for the mitochondrial isoform was 31 ± 3% of the total.

In Vitro Aminoacylation—The full-length mitochondrial and cytoplasmic lysyl-tRNA synthetases were expressed in E. coli and purified. These enzymes were used for aminoacylation assays with gel-purified in vitro expressed transcripts for cytoplasmic and mitochondrial tRNA<sup>lys</sup>. Prior to determining the specific activity of the enzymes, the maximum level of aminoacylation was determined as a percentage for each transcript, using both forms of the enzyme. As calculated by this plateau analysis, approximately 75% of cytoplasmic tRNA<sup>lys</sup> was aminoacylated by the cytoplasmic lysyl-tRNA synthetase. The average maximum levels for aminoacylation obtained using the mitochondrial lysyl-tRNA synthetase were approximately 95% of the cytoplasmic tRNA<sup>lys</sup> transcript and 30% of the mitochondrial tRNA<sup>lys</sup> transcript. Although the two forms of the enzyme aminoacylate both the cytoplasmic and the mitochondrial
tRNA^Lys^aa transcripts, the specific aminoacylation activity with the mitochondrial tRNA^Lys^aa was less than 1% of that with the cytoplasmic tRNA^Lys^aa. For mitochondrial lysyl-tRNA synthetase, the specific aminoacylation activities were 13.3 μmol/min/μmol of protein with cytoplasmic tRNA^Lys^aa and 0.1 μmol/min/μmol of protein with mitochondrial tRNA^Lys^aa. For cytoplasmic lysyl-tRNA synthetase, the specific aminoacylation activities were 3.02 μmol/min/μmol of protein with cytoplasmic tRNA^Lys^aa and 0.02 μmol/min/μmol of protein with mitochondrial tRNA^Lys^aa.

The low aminoacylation activity of the mitochondrial tRNA^Lys^aa transcript was expected, since the mitochondrial tRNA^Lys^aa in vitro transcript has been shown to have a modified hairpin structure due to the lack of a methyl group at position 1 of adenosine 9 (19). The reason for the higher plateau levels of aminoacylation and the 4-fold higher activity of the mitochondrial isoform over the cytoplasmic isoform is not known. This may reflect the difference between the cytoplasmic enzyme expressed with an amino-terminal hexahistidine tag (4) and the mitochondrial form with a carboxyl-terminal tag.

**DISCUSSION**

A variety of strategies are employed that allow a single gene to encode proteins destined for localization in two or more eukaryotic cellular compartments (reviewed in Refs. 20–22). Single genes may give rise to multiple transcripts or single transcripts with multiple translation initiation sites, or the gene transcript may undergo alternative splicings. A single protein may also be localized to different compartments by inefficient targeting (21) or by a chimeric targeting sequence whose activity is modulated by post-translational modification (23).

Where one gene encodes proteins located in both the cytoplasm and mitochondria, the cell frequently generates two types of transcripts from the gene. This can be accomplished by alternate transcription start sites or by alternate splicing of the transcript. In each case, the result is the same. The shorter transcript encodes a nonmitochondrial form of the protein, and the longer transcript encodes a mitochondrial targeting sequence in frame with the protein produced from the initiating Met on the shorter transcript. The translated proteins are identical except for the upstream targeting sequence.

In the human lysyl-tRNA synthetase gene, exon 1 encodes the amino-terminal region of the cytoplasmic form of the enzyme. Exon 2 encodes the amino terminus of the mitochondrial protein, containing the mitochondrial targeting sequence that is presumably cleaved upon import of the protein into the mitochondrion. The third and subsequent exons encode that part of the enzyme that is common to both forms of the synthetase. The first three exons are alternatively spliced solely to incorporate or exclude the second exon that encodes the mitochondrial targeting sequence. This form of alternative splicing, to our knowledge, is thus far unique among enzymes destined for more than one cellular location. This pattern of splicing appears to be conserved among the lysyl-tRNA synthetase genes of humans, mice, zebrafish, and *C. elegans* but not *S. cerevisiae* (Fig. 1B), based on our surveys of the EST and genomic data bases. An exception is *Drosophila*, where although the overall genomic structure is maintained, the mitochondrial transcript is predicted to commence at exon 2. A more detailed analysis of RNA transcripts from these species is needed to confirm these predictions.

As a consequence of the splicing of exon 1 to exon 2, the mRNA transcript encoding the mitochondrial isoform contains a 23-aa upstream open reading frame (uORF), derived largely from exon 1, that terminates 24 nt upstream from the mitochondrial lysyl-tRNA synthetase initiating methionine. The first 20 aa of this uORF are identical to the amino-terminal 20 aa of the cytoplasmic isoform. Such a uORF may have a role in the regulation of the translation of the downstream reading frame (24, 25). The uORF in the mRNA encoding mitochondrial lysyl-tRNA synthetase may be important, since its presence has been conserved among the higher eukaryotes (Fig. 1B). In *Drosophila*, which appears to lack the first exon in the mitochondrial transcript, the second exon encodes a short uORF that is unrelated to the cytoplasmic gene.

According to the scanning model of translation, the ribosome scans the mRNA from the 5′-end and generally initiates translation at the first ATG codon it encounters (26). The translation of the second open reading frame on a transcript can occur by a “leaky scanning” mechanism, where the ribosome does not always recognize the initial ATG if the surrounding sequence is not suitable for efficient initiation. Alternatively, reinitiation of translation can occur after the termination of translation of the uORF (26). The latter mechanism is more likely in the case of mitochondrial lysyl-tRNA synthetase expression, since the sequence around the first ATG (ggagAGTgGc) contains the A at the −3-position and G at the +4-position that characterize a strong initiating sequence (26). Further studies are necessary to determine the mechanism for initiating translation of human mitochondrial lysyl-tRNA synthetase.

A number of human tRNA synthetases have been described in both their cytoplasmic and mitochondrial forms. In most cases, two genes give rise to the separate forms of each enzyme, e.g. histidyl tRNA synthetase (8), phenylalanyl-tRNA synthetase (27, 28), and tryptophanyl-tRNA synthetase (29, 30). However, in the case of glycyl-tRNA synthetase, there are two proteins derived from a single gene (10). This is due to translation from alternate initiation codons resulting in inclusion or exclusion of a mitochondrial targeting sequence (10). Although other organisms contain numerous examples of cytoplasmic and mitochondrial proteins generated by alternate splicing, human lysyl-tRNA synthetase appears to be unique, or at least unusual, in that the mitochondrial targeting sequence is encoded on the second exon. More commonly, alternate transcription initiation provides alternative 5′ exons, with the mitochondrial targeting sequence being on the first exon e.g. human dUTPase (31). An examination of GenBank™ data bases leads us to conclude that only human glycyll- and lysyl-tRNA synthetases are likely to have both cytoplasmic and mitochondrial isoforms encoded by single genes.²

Although between 15 and 30% of the mitochondrial tRNA^Lys^aa transcript could be aminoacylated by high levels of enzyme, the specific activities for aminoacylation of the mitochondrial tRNA^Lys^aa with both mitochondrial and cytoplasmic enzymes were less than 1% of those obtained with the cytoplasmic tRNA^Lys^aa. This was probably due to the incorrect folding of the *in vitro* mitochondrial tRNA^Lys^aa transcript into a modified hairpin structure because of the lack of a methyl group, at position 1 of adenosine 9, that is found in the native tRNA (19). The m^1^A^9^ modification appears to be necessary for the proper folding of the wild-type tRNA^Lys^aa (19, 32).

In a tRNA, the “discriminator base” at position 73 in the acceptor stem is often important in ensuring the specific aminoacylation of the tRNA (33). In most cases, synthetases do not efficiently aminoacylate cognate tRNAs with substitutions at position 73. However, the cytoplasmic isoform of the lysyl-tRNA synthetase is tolerant of variations at the G^73^ discriminator base position of cytoplasmic tRNA^Lys^aa and can consequently aminoacylate the *E. coli* tRNA^Lys^aa that has A at position 73 (4). The data presented here show that essentially the same

²A. Palmitessa, M. P. King, and E. Davidson, unpublished observations.
protein is normally required to aminoacylate mitochondrial tRNA\(^{Lys}\) with an A at the discriminator position (34) and cytoplasmic tRNA\(^{Lys}\) that has a G at the discriminator position. In contrast, the human glycyll-tRNA synthetase, which is encoded by a single gene and functions in both the cytoplasm and mitochondria, is unable to aminoacylate E. coli tRNA\(^{Gly}\) (9) and shows specific requirements for the nucleotide at the discriminator position (35). This reflects the fact that both cytoplasmic and mitochondrial tRNA\(^{Gly}\) have the same nucleotide at position 73.

One gene appears to encode both cytoplasmic and mitochondrial forms of S. cerevisiae cysteinyl-tRNA\(^2\), histidyl-tRNA (36, 37), and valyl-tRNA synthetases. Of those cases where one gene provides both cytoplasmic and mitochondrial tRNA synthetases, S. cerevisiae cytoplasmic and mitochondrial tRNA\(^{Cyts}\) also conserve the discriminator base \((U\^{73})\), while S. cerevisiae cytoplasmic and mitochondrial tRNA\(^{His}\) and tRNA\(^{Val}\) differ at the discriminator position of these tRNAs. In the case of yeast tRNA\(^{Val}\), there are several cytoplasmic tRNAs, one of which has a different discriminator nucleotide from that common to the mitochondrial tRNA\(^{Val}\) and other cytoplasmic tRNA\(^{Val}\).

The human lysyl-tRNA synthetase also interacts with the different secondary and tertiary structures of cytoplasmic and mitochondrial tRNA\(^{Lys}\). The mitochondrial tRNA\(^{Lys}\) has a relatively small D loop, 5 nucleotides shorter than that of cytoplasmic tRNA\(^{Lys}\), and has also been suggested to have a different tertiary conformation (38). The mitochondrial tRNA\(^{Lys}\) has a much larger interstem angle (\(-140^\circ\)) between the anticodon and acceptor stems than that of the cytoplasmic tRNA\(^{Lys}\) (80–90\(^\circ\)) such that the overall shape of the tRNA does not conform to the canonical L shape (38). This suggests that the human lysyl-tRNA synthetase has considerable flexibility in regard to substrate tRNAs.

A number of mutations in the human mitochondrial tRNA\(^{Lys}\) are associated with human diseases (39, 40). Human cell lines containing the A834GG mutation have decreased levels of aminoacylated mitochondrial tRNA\(^{Lys}\) (41). The identification of human mitochondrial lysyl-tRNA synthetase will allow investigation into whether pathogenic mutations in mitochondrial tRNA\(^{Lys}\) affect interactions with lysyl-tRNA synthetase and show whether defects in these interactions contribute to the mechanism of pathogenesis.

Acknowledgments—We thank Dr. Kiyohtaka Shiba for the gift of plasmids pM386 and pLysF119, Dr. Ya-Ming Hou for a critical reading of the manuscript, and Dr. Ya-Ming Hou and Dr. Rich Lipman for technical advice on the aminoacylation assays.

REFERENCES
1. Onesti, S., Miller, A. D., and Brick, P. (1995) Structure 3, 163–176
2. Cusack, S., Yaremchuk, A., and Tukalo, M. (1996) EMBO J. 15, 6321–6334
3. Ibbi, M., Morgan, S., Curnow, A. W., Pridmore, D. R., Vothknecht, U. C., Gardner, W., Lin, W., Woese, C. R., and Soll, D. (1997) Science 278, 1119–1122
4. Shiba, K., Stello, T., Motejie, H., Noda, T., Musier-Forsyth, K., and Shimmel, P. (1997) J. Biol. Chem. 272, 22869–22816
5. Quevillon, S., Robinson, J. C., Berthonneau, E., Siatecka, M., and Mirande, M. (1999) J. Mol. Biol. 285, 183–195
6. Rho, S. B., Lee, K. H., Kim, J. W., Shiba, K., Jo, Y. J., and Kim, S. (1996) Proc. Natl. Acad. Sci. U. S. A. 93, 10128–10133
7. Nichols, R. C., Blinder, J., Pai, S. I., Ge, Q., Targoff, I. N., Plotz, P. H., and Liu, P. (1996) Genomics 36, 210–213
8. O’Hanlon, T. P., Raben, N., and Miller, F. W. (1995) Biochem. Cell Biol. 210, 556–566
9. Shiba K., Schimmel P., Motejie H., and Noda T. (1994) J. Biol. Chem. 269, 20954–20957
10. Mudge S. J., Williams J. H., Eyre H. J., Sutherland G. R., Cowan P. J., and Power D. A. (1998) Gene (Amst.) 209, 45–50
11. Mirande, M., Le Corre, D., Riva, M., and Waller, J. P. (1986) Biochimie (Paris) 68, 1091–1107
12. Gatti D. L., and Tzagoloff A. (1991) J. Mol. Biol. 218, 557–568
13. Varshavsky, A. (1996) Proc. Natl. Acad. Sci. U. S. A. 93, 12142–12149
14. Frohman, M. A. (1993) Methods Enzymol. 218, 340–356
15. Sampson, J. R., and Uhlenbeck, O. C. (1998) Proc. Natl. Acad. Sci. U. S. A. 85, 1033–1037
16. Fechter, P., Rudinger, J., Giege, R., and Theobald-Dietrich, A. (1998) FEBS Lett. 436, 99–103
17. Claros M. G. (1995) Comput. Appl. Biosci. 11, 441–447
18. Li, B., Oestreich, S., and de Lange, T. (2000) Cell 101, 471–483
19. Helm, M., Brule, H., Degoul, F., Cepanez, C., Leroux, J. P., Giege, R., and Florentz, C. (1998) Nucleic Acids Res. 26, 1636–1643
20. Martin, N. C., and Hopper, A. K. (1994) Biochimie (Paris) 76, 1161–1167
21. Danpure, C. (1997) Trends. Cell. Biol. 5, 230–238
22. Small, I., Wintz, H., Akashi, K., and Mireau, H. (1998) Plant Mol. Biol. 38, 265–277
23. Anandatheerthavarada, H. K., Biswas, G., Mullick, J., Sepuri, N. B., Otvos, L., Pain, D., and Avadhanl, N. G. (1999) EMBO J. 18, 5842–5848
24. Geballe, A. P. (1996) in Translational Control (Hershey, J. W. B., Mathews, M. B., and Sonenberg, N., eds) pp. 173–197, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
25. Gray, N. K., and Wickens, M. (1998) Annu. Rev. Cell Dev. Biol. 14, 399–458
26. Kozak, M. (1999) Gene (Amst.) 234, 187–208
27. Buillard, J. M., Cai, Y. C., Demeler, B., and Spremulli, L. L. (1999) J. Mol. Biol. 288, 567–577
28. Rodova, M., Anklilova, V., and Safro, M. G. (1999) Biochem. Cell Biol. 253, 655–773
29. Frolova, L., Sudomoina, M. A., Grigoriev, A., Zinovieva, O. L., and Kisselev, L. L. (1991) Gene (Amst.) 109, 291–296
30. Jorgensen, R., Sogaard, T. M., Rossing, A. B., Martensen, P. M., and Justesen, J. (2000) J. Biol. Chem. 275, 16820–16826
31. Ladner, R. D., and Caradonna, S. J. (1997) J. Biol. Chem. 272, 19072–19080
32. Helm, M., Giege, R., and Florentz, C. (1999) Biochemistry 38, 13338–13346
33. Lee, C. P., Mandal, N., Dynon, M. R., and RajBhandary, U. L. (1993) Proc. Natl. Acad. Sci. U. S. A. 90, 7149–7152
34. Sprinzl, M., Horn, H., Brown, M., Ioudovitch, A., and Steinberg, S. (1998) Nucleic Acids Res. 26, 148–153
35. Hippe, D., Shiba, K., Henderson, B., and Schimmel, P. (1995) Proc. Natl. Acad. Sci. U. S. A. 92, 5550–5552
36. Chatton, B., Walter, P., Ebel, J. P., Lacroute, F., and Fasiolo, F. (1988) J. Biol. Chem. 263, 52–57
37. Natsoulis, G., Hilper, F., and Fink, G. R. (1986) Cell 46, 235–245
38. Leehey, M. A., Squassoni, C. A., Friederich, M. W., Mills, J. B., and Hagerman, P. J. (1995) Biochemistry 34, 16235–16239
39. Masucci, J., Davidson, M., Koga, Y., Schon, E. A., and King, M. P. (1995) Mol. Cell. Biol. 15, 2872–2881
40. Davidson, E., and King, M. P. (1997) Trends. Cardiovasc. Med. 7, 16–24
41. Enriquez, J. A., Chomyn, A., and Attardi, G. (1995) Nat. Genet. 10, 47–55