Mouse pseudouridine synthase 1: gene structure and alternative splicing of pre-mRNA

Jianguang CHEN and Jeffrey R. PATTON[1]

Department of Pathology, School of Medicine, University of South Carolina, Columbia, SC 29208, U.S.A.

Evidence for the alternative splicing of the message for mouse pseudouridine synthase 1 (mPus1p) was found when several expressed sequence tag clones were completely sequenced. The genomic DNA for the *MPUS*1 gene (6.9 kb) was cloned from a mouse genomic library; the gene contains seven exons, of which three are alternatively spliced. In addition, one of the internal exons (exon VI) is unusually large. RNase protection analysis confirmed that several alternatively spliced messages were present in mouse tissues and cells in culture. A Western blot of total cellular protein from mouse tissues and cultured cells was reacted with an antibody specific for mPus1p; at least three proteins were detected. One protein corresponds to the predicted molecular mass of mPus1p (44 kDa) and is the most abundant. The two other isoforms, one 2 kDa larger and one 7 kDa smaller than mPus1p, were differentially expressed. The cDNA species for the three isoforms were cloned into expression plasmids; the proteins were synthesized *in vitro* and tested for pseudouridine synthase activity. The two isoforms, one containing an insert of 18 amino acids in a region of the enzyme assumed to be critical for activity, and the other, which has a deletion of the protein coding potential of two exons, were both inactive on tRNA substrates that mPus1p modifies.

Key words: expressed sequence tag, genomic DNA, *in vitro* translation tRNA, truA.

INTRODUCTION

Pseudouridine is an abundant modification found in stable RNA species including tRNA [1], rRNA [2] and small nuclear RNA [3]. Pseudouridine is a post-transcriptional modification [4] and pseudouridine synthases have been isolated from prokaryotes and eukaryotes [5–18]. The enzymes are organized into families [19], with the truA family having the most members. This family includes pseudouridine synthase I (PSUI; truA [10]) from *Escherichia coli*, which modifies positions 38, 39 and 40 of tRNA; Pus1p and mPus1p from yeast [14] and mouse [7] respectively, which modify positions 27 and 28 and the anticodon of intron-containing tRNA species; and yeast pseudouridine synthase 3 (deg1 [12]), which modifies positions 38 and 39 in yeast tRNA species. Yeast pseudouridine synthase 1 (Pus1p) has been shown to modify additional positions on tRNA *in vivo* [20] and U2 small nuclear RNA *in vivo* and *in vitro* [21]. Deletion of pseudouridine synthase genes in *E. coli* and yeast can result in considerably slower growth rates for the deleted strains [12,13,22,23].

When the cloning and characterization of mouse pseudouridine synthase 1 (mPus1p) was reported it was noted that the message for the protein was probably alternatively spliced [7]. Alternative splicing of pre-mRNA is an effective way for cells to generate diversity in the repertoire of proteins found in the cell and also a way of regulating the expression of genes [24–26]. For mPus1p, alternatively spliced mRNA species could code for protein isoforms with different substrate specificities.

Here we report the genomic structure for the *MPUS*1 gene and show the pathways for alternative splicing of several isoforms. We give evidence that at least two of these alternatively spliced mRNA species are translated *in vivo*. Two of these isoforms were cloned into expression vectors; the proteins were translated *in vitro* and tested in pseudouridine synthase assays with known substrates of mPus1p. We find that either change in the primary sequence of mPus1p results in a loss of pseudouridine synthase activity *in vitro*.

EXPERIMENTAL

Materials

I.M.A.G.E. Consortium (LLNL) cDNA clones (accession numbers AA017902 and AA517185 [27]) were purchased from Genome Systems and the inserts were completely sequenced [28]. The published sequences of other expressed sequence tags (ESTs) (accession numbers AU051076, AI613887 and AU035787) were used directly for comparison without sequencing. We used the Baylor College of Medicine website (http://www.hgsc.bcm.tmc.edu/SearchLauncher/) to identify the clones with the sequenced cDNA for mPus1p [7] for comparison. The yeast $tRNA^{Phe}$ clone was obtained from Dr O. Uhlenbeck [29]; the clones of yeast $tRNA^{Val}$ (with Met anticodon UAC) and the yeast $tRNA^{Ile}$ with intron were obtained from Dr F. Fasiolo [12,30]. The polyclonal antibody against recombinant mPus1p was raised in rabbits (Alpha Diagnostic International, San Antonio, TX, U.S.A.). The antibody was purified by incubating the serum with nitrocellulose-bound antigen (recombinant mPus1p [7]) and eluting it at low pH [31].

Cloning and sequencing of genomic DNA

A mouse (129 SvJ) Lambda phage genomic library (Stratagene, La Jolla, CA, U.S.A.) was screened with the insert from clone AA517185 and several overlapping lambda clones were identified

Abbreviations used: EST, expressed sequence tag; mPus1p, mouse pseudouridine synthase 1; ORF, open reading frame; PSUI, pseudouridine synthase I; Pus1p, yeast pseudouridine synthase 1; RPA, ribonuclease protection assay; RT–PCR, reverse-transcriptase-mediated PCR; UTR, untranslated region.

[1] To whom correspondence should be addressed (e-mail patton@med.sc.edu).

and plaque-purified. Subclones that covered the entire gene were generated from the phage clones and the ends were sequenced [28]. Any gaps in the sequence were closed by generating primers that were used to generate PCR fragments [32] from the original lambda clones or subclones. These fragments were in turn cloned and sequenced. In addition, a series of nested deletions of some of the genomic subclones were generated and subsequently sequenced [31]. The sequence, from −2419 bp relative to the start of transcription to 308 bp past the polyadenylation site, was determined on both strands or the same strand twice [28].

Recombinant protein expression

Fragments containing the coding regions for the deletion of exons III and IV (mPus1p$^{\Delta Ex3+4}$) and the insertion of exon IV (mPus1p^{alt}) were generated from plasmids AA017902 and AA517185 respectively by using the method and primers described in detail for the expression of mPus1p [7]. The fragment generated by PCR was inserted into pGEMT (Promega) and sequenced [28]. The *Nde*I–*Bam*HI fragments from pGEMT-mPus1^{alt} or pGEMT-mPus1p$^{\Delta Ex3+4}$ were inserted into pET16b (Novagen) cut with *Nde*I and *Bam*HI. The open reading frame (ORF) included the additional amino acid sequence MGHH-HHHHHHHHSSGHIEGRH (single-letter amino acid codes) at the N-terminus of the mPus1p coding sequences. The DNA from the expression clones was used directly in the T7 coupled transcription/translation reaction (TnT® T7 Coupled Reticulocyte Lysate System; Promega) to produce unlabelled protein *in vitro* in accordance with the manufacturer's recommendations.

Western blots

Samples of total protein from mouse tissues or cultured RAW264.7 cells (120 μg) or aliquots from the TnT reactions (5 μl of a 50 μl reaction) were subjected to PAGE [10 % (w/v) gel] [33] and transferred to nitrocellulose. Purified rabbit anti-mPus1p polyclonal antibody was used as the primary antibody and the secondary antibody was horseradish-peroxidase-conjugated, donkey anti-rabbit IgG whole antibody (Amersham/Pharmacia Biotech). The enhanced chemiluminescence (ECL®) Western blotting detection reagents (Amersham/Pharmacia Biotech) were used to detect antibody reactivity with XAR film (Kodak).

Transcription *in vitro*

The plasmid DNA species containing the tRNA genes were digested with *Bst*N1; all were transcribed with T7 RNA polymerase as described [34] in the presence of 50 μCi (800 Ci/mmol) [α-^{32}P]UTP with a final concentration of UTP of 50 μM. The other, unlabelled, NTPs were at 500 μM. The RNA species were purified on 10 % (w/v) polyacrylamide/8.3 M urea gels before use in assays.

Primer extension, ribonuclease protection assay (RPA) analysis and reverse-transcriptase-mediated PCR (RT–PCR)

Total mouse RNA was prepared from mouse tissues and RAW264.7 cells (mouse monocyte–macrophage line) as described [35], hybridized to the ^{32}P-end-labelled [31] gel-purified primer (5′-CCCAGACCCAGCCACCTCGG-3′). The primer was extended with Moloney-murine-leukaemia virus reverse transcriptase (100 units; Promega) at 37 °C for 60 min as described [31]. The products were subjected to electrophoresis on a 7 % (w/v) polyacrylamide/8.3 M urea gel; the gel was then dried and exposed to film. This primer hybridizes to a sequence within exon II, between nt 67 and 86 from the start of exon II.

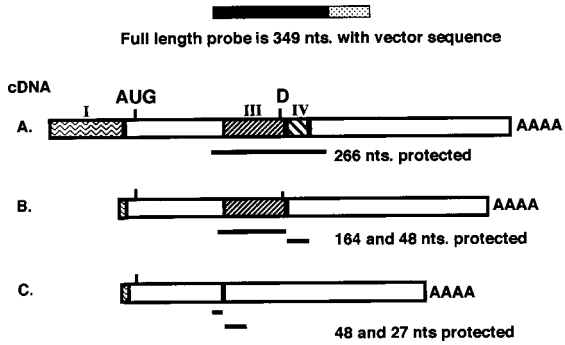


Figure 1 mRNA for mPus1p is alternatively spliced

Diagram of three cDNA species (denoted A, B and C) representing alternatively spliced forms of the mRNA for mPus1p that were derived from the sequences of EST clones (see the Experimental section). The lengths of the protected portion(s) of an anti-sense probe used in RPAs are given below each cDNA. A diagram of the full-length probe, including the vector sequence (shaded), is shown at the top. The roman numerals refer to exons on the *MPUS*1 gene (Figure 2, upper panel).

The RNA species were also used in RPAs (RPA II®; Ambion) with the probe shown in Figure 1. The probe was made by generating an RT–PCR fragment [32,36] from poly(A)$^+$ RNA (1 μg) from mouse kidney with primers in exons II and V (forward primer, 5′-CAAAGCGAAAAATTGTGCTG-3′; reverse primer, 5′-ATCAGCCACACCTTTAGGGA-3′) and cloning the fragment into pGEMT (Promega). The parameters for PCR are listed below. Anti-sense RNA was synthesized from *Ava*I-cut DNA with SP6 RNA polymerase and [^{32}P]UTP as listed above for the generation of tRNA substrates. The RPA products were subjected to electrophoresis on a 10 % (w/v) polyacrylamide/8.3 M urea gel; the gel was then dried and exposed to X-ray film. Quantification of the levels of the products was accomplished with a PhosphorImager and either ImageQuan 3.0 (Molecular Dynamics) or Quantity One software (Bio-Rad).

The RT–PCR assay shown in Figure 5 used the same primers as those listed for the generation of the RPA probe but, instead of poly(A)$^+$ RNA, 10 μg of total RNA from the mouse tissues and cultured cells was used [32,36]. The PCR parameters were 94 °C for 5 min, then 35 cycles of 94 °C for 30 s, 55 °C for 30 s and 72 °C for 1 min. A final incubation of 20 min at 72 °C was performed after the cycling was complete.

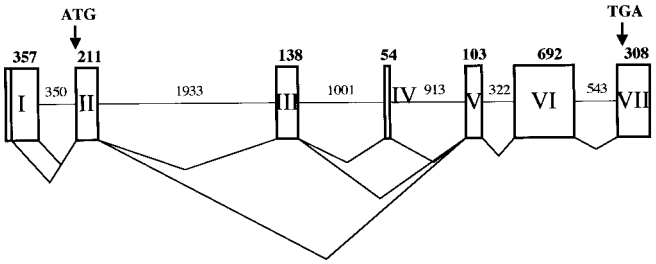
Assays for pseudouridine formation

The reactions with the pseudouridine synthases were performed in 100 mM NH_4Cl/10 mM dithiothreitol/50 mM Tris/HCl (pH 7.5)/2 mM $MgCl_2$. The RNA substrates, without enzyme, extract or dithiothreitol, were heated to 65 °C for 2 min and left to cool slowly to 35 °C (this took 30–40 min). The amount of mPus1p or isoform translated *in vitro* was one-tenth of a standard TnT® reaction (5 μl), estimated from the Western blots to be approx. 10 ng of synthase. After addition of the enzyme, the reactions were incubated for 20 min. To determine the amount of pseudouridine formation, the ^{32}P-labelled tRNA species were isolated from the reaction, digested with nuclease P1 and chromatographed on TLC plates (cellulose; Kodak) with propan-2-ol:HCl:water (70:15:15, by vol.) as the mobile phase [37]. The spots were identified by autoradiography and quantified by scraping the spot from the plate and counting by liquid scintillation.

RESULTS

Evidence for alternative splicing found in ESTs

In a previous study [7], bands of approx. 1.4, 1.6 and 4.3 kb were observed in a Northern blot of total RNA from mouse tissues or cells in culture probed with the cDNA for mPus1p. The nucleic acid sequence of mPus1p cDNA was used to identify EST clones that might be the result of alternative splicing of the mPus1p pre-mRNA. Several of these were sequenced and three iterations of the sequence were found. One was basically the same as that already seen with mPus1p (Figure 1, cDNA B), in other words the ORF would give a protein with an amino acid sequence identical with the published sequence for mPus1p [7]. Another form had additional sequence on the 5′ end and an insert of 54 bp within the sequence (Figure 1, cDNA A). There was an in-frame start codon within the additional 5′ sequence that could add 30 amino acid residues to the N-terminus of the protein if it were used. The additional 54 bp was also in frame and would insert 18 residues into a highly conserved region of mPus1p, just after the aspartate residue known to be essential for activity in *E. coli* PSUI [38,39]. Finally, there was another form with no additional sequence at the 5′ end (Figure 1, cDNA C) but lacking both the region encoding the essential aspartate residue and the additional 54 bp (marked III and IV in Figure 1). The sequence after the deletion was still in frame, so the amino acid sequence was the same as mPus1p after that point, but it was lacking a region of the protein known to be essential for activity in PUSI [38]. Additional EST clones were found that extended the 5′ end by a few nucleotides but the types of spliced isoform did not vary from the three presented above.



5′ splice site 3′ splice site

```
Exon Ia    CC|GTGAGT
Exon Ib    GG|GTAAAG...TGTTTTTCTCAG|CC Exon II
Exon II    AG|GTGAGG...TTTTTTCTTTAG|AG Exon III
Exon III   AG|GTATGT...CTGTCCCCGTAG|CC Exon IV
Exon IV    GG|GTAAGG...TTTTCTTTGCAG|GG Exon V
Exon V     GG|GTAAGC...CCTGTCCCCCAG|GA Exon VI
Exon VI    AG|GTAGGA...CCTTTCCTACAG|GT Exon VII

Consensus ag|GTaagt...yyyyyyyynyAG|gt
```

Figure 2 Gene structure and splice junction sequences of the MPUS1 gene

Upper panel: the exons (large blocks marked with roman numerals) and introns (thin lines between the exons) are illustrated with the sizes in bp listed above each feature. The start and stop codons of mPus1p are marked in exons II and VII respectively. The modes of known alternative and constitutive splicing are indicated below the gene by lines joining the exons. Lower panel: sequences of the splice site junctions for all the exons. The consensus 5′ and 3′ splice site sequences are shown for comparison at the bottom of the panel. The presence of capital letters in the consensus sequence indicates absolute conservation; a lowercase letter indicates a predominance of that nucleotide.

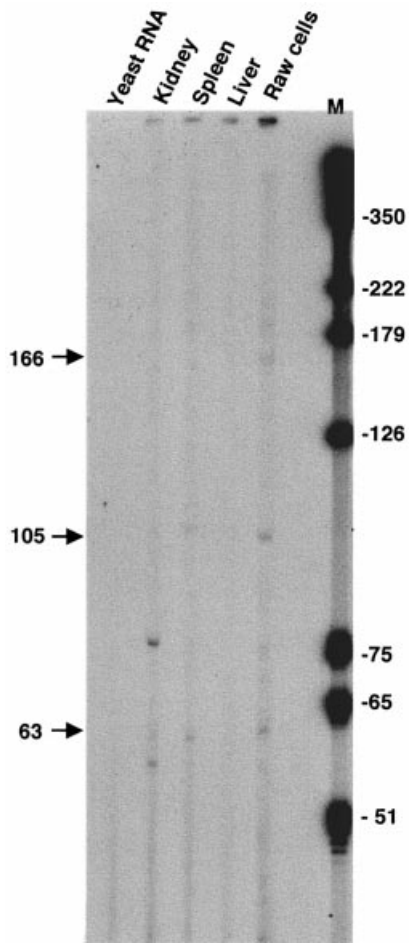


Figure 3 Primer extension analysis of total RNA species from mouse tissues and cultured cells

Total RNA species (10 μg) from mouse tissues and RAW 264.7 cells were hybridized to a ^{32}P-labelled end-labelled primer and extended with reverse transcriptase. The products were subjected to electrophoresis on a 7% (w/v) polyacrylamide/8.3 M urea gel; an autoradiograph of a portion of the gel is shown. The sizes (in nucleotides) of the DNA markers (lane M) are given at the right. The positions of three of the extension products are indicated at the left, in nucleotides.

MPUS1 genomic sequence

To understand better the pathways for the alternative splicing of the mPus1p pre-mRNA, the genomic sequence for the *MPUS*1 gene was determined (AF269250; see the Experimental section). The gene structure of seven exons and six introns is shown in Figure 2 (upper panel). The gene is 6.9 kb long and the 5′ end of the first exon was determined by primer extension on mouse total RNA from tissues and cultured RAW 264.7 cells (see the Experimental section). The results of the extension of a primer located in exon II are shown in Figure 3. Weak bands appeared in the lanes where kidney, spleen and RAW 264.7 cells (a mouse macrophage cell line) total RNA was included in the reaction. The products observed in the lane with the RNA from RAW cells are indicated at the left. If an mRNA contained only the first part of exon I, the product at 166 nt would indicate a 5′ end for the gene that was extended 45 nt past the 5′ end of a putative full-length cDNA for mpus1 (AB041563). This would make the size

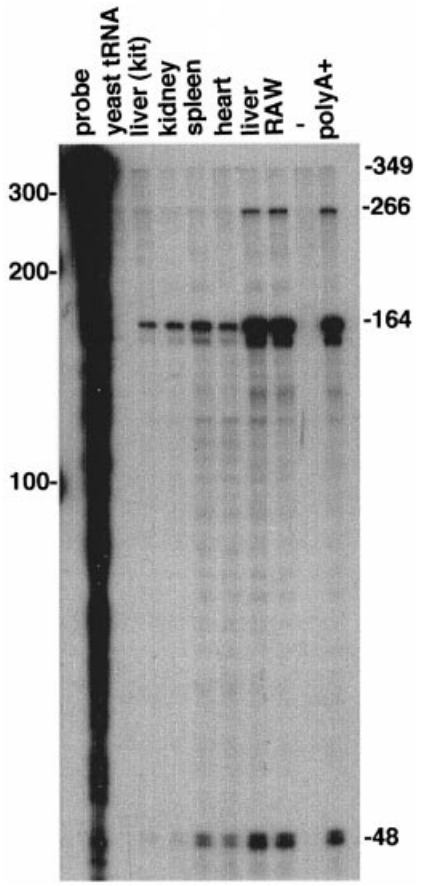


Figure 4 RPA of RNA from tissues and RAW 264.7 cells

Autoradiograph of a 10% (w/v) polyacrylamide/8.3 M urea gel of the products of the RPA analysis. The types of RNA used in the hybridization with the probe are shown at the top. All of the RNA species were total RNA (5–10 μg) except the poly(A)$^+$ lane (1 μg), which was isolated from mouse kidney. Yeast tRNA indicates that yeast RNA was used in the hybridization. RAW denotes total RNA from RAW264.7 cells and a dash indicates no RNA was used in the hybridization. The sizes of RNA markers (in nucleotides) are indicated at the left; the sizes of the full-length probe and protected products in nucleotides are listed (in nucleotides) at the right.

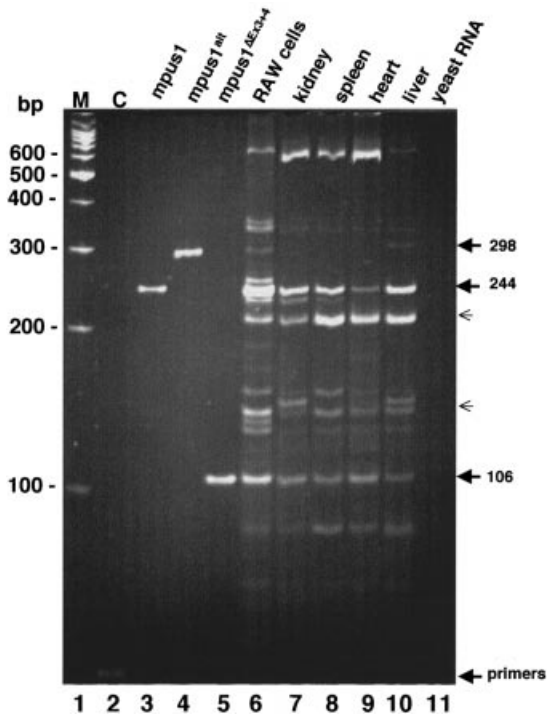


Figure 5 mRNA corresponding to the deletion of exons III and IV is present in total RNA from mouse tissues and cells in culture

Shown is an ethidium-bromide-stained 5% (w/v) polyacrylamide, non-denaturing gel of PCR and RT–PCR reactions. In lanes 6–10 total RNA from mouse tissues or cells was used in RT–PCR as described in the Experimental section; in lane 11 yeast RNA was used instead of mouse RNA. The expression plasmids pET16b-mPus1p (lane 3), pET16b-mPus1p^{alt} (lane 4) and pET16b-mPus1p$^{\Delta Ex3+4}$ (lane 5) were used as templates for PCR and served as markers for the expected amplified fragment size. A control (C) with no added DNA or RNA is shown in lane 2; DNA size markers (M) are shown in lane 1, with sizes indicated at the left. The positions of the three expected fragments (bold arrows) are indicated at the right and the position of the primers is shown at the bottom. The thin arrows at the right indicate additional bands that were cloned and sequenced (see the text).

of exon I 402 bp. The other extension products could result from the actual 5′ ends of other mRNA species, strong stops to reverse transcriptase due to secondary structure, or degraded mRNA species. Longer extension products that would result from mRNA species containing the entire exon I were not observed.

The 3′ end of exon VII is the polyadenylation site derived from the cDNA sequences and is 16 bp downstream of the end of a polyadenylation signal sequence ATTAAA [40]. The resulting 3′ unstranslated region (UTR) is 260 nt long. Exon VI is a relatively large non-terminal exon (692 bp), an exception to the under 300 nt size limit for internal exons that participate in the exon-definition mode of splicing [41,42]. The introns that flank exon VI are relatively small (322 and 543 nt; Figure 2, upper panel), which should promote the inclusion of that large internal exon, given the results of Sterner et al. [43].

Exon I can be included in its entirety and results in an mRNA with an extended 5′ end (represented by cDNA A in Figure 1). Alternatively, just the first 35 nt could be linked to exon II to give the shorter 5′ ends seen with the cDNA species B and C in Figure 1. The 35 nt is based on the known ends of the cDNA species that have been sequenced. The size of this portion of exon I might in fact be closer to 80 nt, given the results of the primer extension analysis (see above). Either 5′ splice site sequence was a good match to the consensus (Figure 2, lower panel) but the 3′ splice site just upstream of exon II contained a purine in the stretch of pyrimidines (see Figure 2, lower panel). Depending on whether just the first 35 nt or the entire exon I is included in the mRNA, the 5′ UTR will be either 45 or 367 nt when the AUG in exon II is used (see Figure 1). If the alternative AUG found in exon I is used then the 5′ UTR will be 277 nt.

The additional 54 bp found in cDNA A (Figure 1) is simply the inclusion of exon IV in the mRNA. There is a G in the polypyrimidine stretch at the 3′ splice site preceding exon IV, whereas the 3′ splice site preceding exon V is a perfect match with the consensus 3′ splice site sequence (see Figure 2, lower panel). Therefore the skipping of exon IV would be favoured over inclusion owing to the relative weakness of the 3′ splice site preceding exon IV.

Both exon III and exon IV are skipped in the generation of cDNA C (Figure 1). The 3′ splice site before and the 5′ splice site after exon III are both excellent matches to the consensus splice site sequences (Figure 2, lower panel). In addition, at 138 nt, exon III is greater than the minimum length observed for internal exons [42]. Therefore the skipping of exon III seems to be a regulated process (see results below).

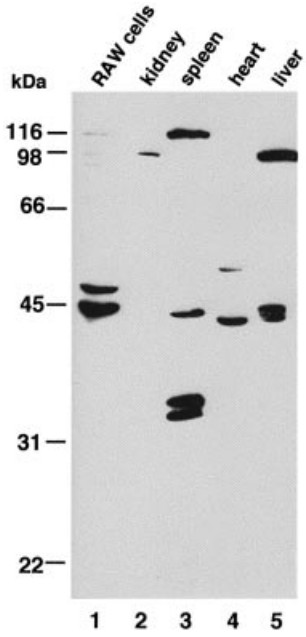


Figure 6 Western blot of protein from mouse tissues and RAW 264.7 cells

Total protein samples (120 μg) were subjected to SDS/PAGE [10% (w/v) gel], transferred to nitrocellulose and reacted with the purified rabbit anti-mPus1p polyclonal antibody (primary; see the Experimental section) and then a horseradish-peroxidase-conjugated, donkey anti-rabbit IgG whole antibody (secondary). The chemiluminesence was detected on XAR film (1 h exposure; see the Experimental section). The positions of molecular mass markers are indicated at the left.

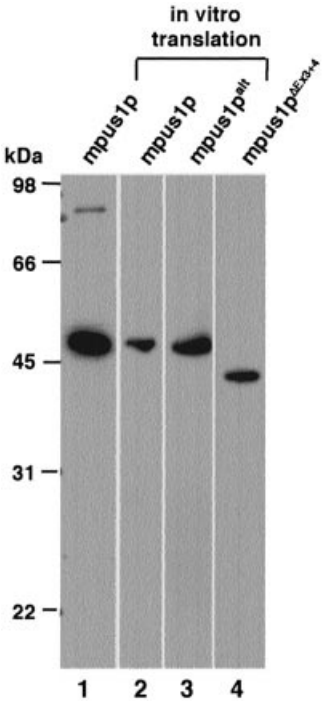


Figure 7 Western blot of recombinant mPus1p and isoforms translated *in vitro*

Purified mPus1p (lane 1) or aliquots of the transcription–translation reactions *in vitro* with pET16b-mPus1p (lane 2), pET16b-mPus1p^{alt} (lane 3) and pET16b-musp1p$^{\Delta Ex3+4}$ (lane 4) DNA were subjected to SDS/PAGE [10% (w/v) gel] [33]. The proteins were transferred to nitrocellulose, reacted with the purified rabbit anti-mPus1p polyclonal antibody (primary; see the Experimental section) and then a horseradish-peroxidase-conjugated, donkey anti-rabbit IgG whole antibody (secondary). The chemiluminesence was detected on XAR film (18 min exposure; see the Experimental section); the positions of molecular mass markers are indicated at the left.

Expression of mPus1p isoforms *in vivo*

An RPA was used to show that RNA species for these isoforms were present in total RNA from mouse tissues and RAW 264.7 cells. An anti-sense probe spanning the region of alternative splicing is shown at the top of Figure 1. Figure 4 shows that in all the samples there were bands at 164 and 48 nt that corresponded to the products expected for the mRNA represented by cDNA B in Figure 1. These were also the most abundant products found in any of the samples. The probe would have to loop out where exons III and V are brought together; the additional bands near 164 and 48 nt were probably due to the generation of heterogeneous ends on the protected fragments during the RNase digestion step in the RPA method. The protected product at 266 nt, corresponding to an mRNA containing the additional 54 nt, was found in most of the samples, although at very low levels in some (Figure 4).

Evidence for the presence of mRNA with the deletion of exons III and IV was found when the relative intensities at 164 and 48 nt were compared with a PhosphorImager. For example, in spleen there was significant molar excess of the 48 nt protected fragment over what would be predicted if only the mRNA represented by cDNA B were present in the sample. In addition, evidence for this RNA was seen in the results of RT–PCR on total RNA from tissues and RAW 264.7 cells (Figure 5). When aliquots of total RNA were used in RT–PCR reactions, a PCR product diagnostic of the presence of the mRNA corresponding to cDNA C (106 bp) was present in all the samples except yeast RNA (see Figure 5). The expected products for the other two isoforms (a 298 bp product for the mRNA corresponding to cDNA A and a 244 bp product for the mRNA corresponding to cDNA B) were seen in the RAW cell sample. However, the 298 bp product was not detected in the tissue samples under these PCR conditions. Figure 5 (lanes 3–5) shows the sizes of the PCR products from the three expression plasmids and serve as markers for the RT–PCR assays. The identities of the bands in the RT–PCR lanes were confirmed by elution of the fragments from a gel, cloning them into the pGEMT vector (see the Experimental section) and then sequencing the clones. The thin arrows mark bands that were cloned and sequenced but matched a mouse virus-like (VL30) retrotransposon BVL-1 [44], mouse very-long-chain acyl-CoA synthetase [45] or a fatty acid transporter [46].

To determine whether these RNA species were being used as templates for translation we used a rabbit polyclonal anti-mPus1p antibody that was purified against nitrocellulose-bound recombinant mPus1p (see the Experimental section for details) to probe a Western blot of total protein from mouse tissues and RAW264.7 cells. In RAW 264.7 cells (Figure 6) two bands of approx. 44 and 46 kDa were observed. The more abundant band at 44 kDa corresponded to the predicted molecular mass of mPus1p of 44098 Da. The minor band at 46 kDa corresponded to the predicted size of a protein containing the additional 18 amino acid residues encoded by the 54 nt (46243 Da) but not to the added N-terminal 30 residues from the extended 5′ sequence (47507 Da) or a protein containing both of these regions (49652 Da). In lane 3 with protein from the spleen, there were two bands at approx. 36 and 37 kDa in addition to a band at

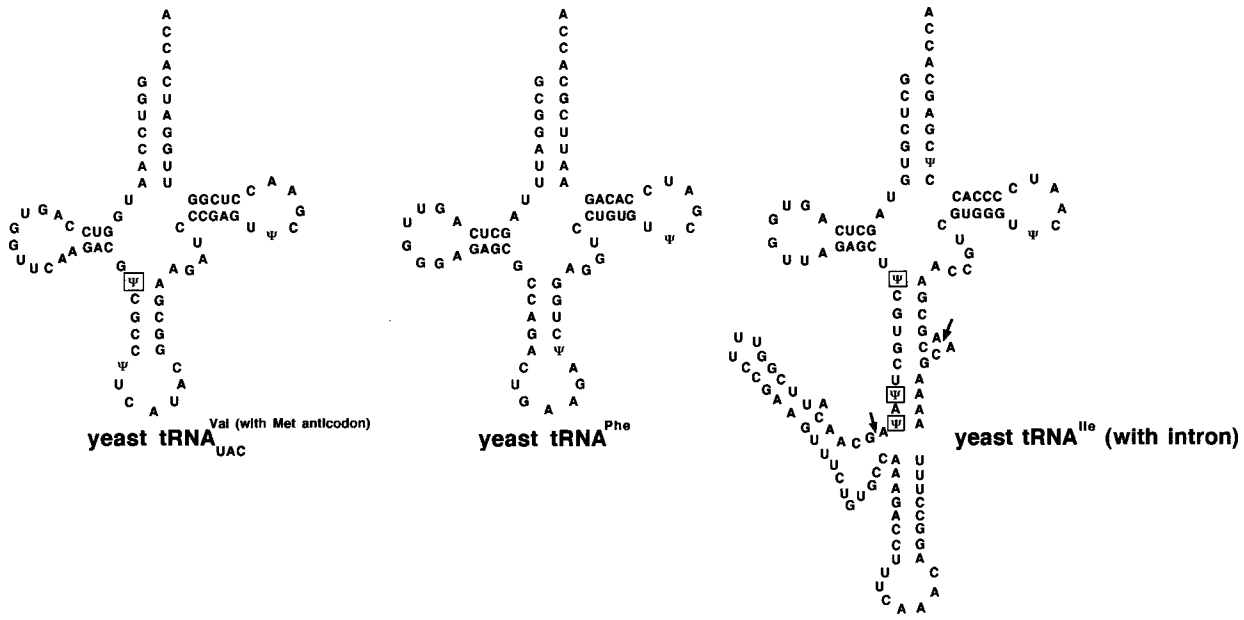


Figure 8 Primary and secondary structures of the three tRNA substrates used in the experiments

These are the sequences of the tRNA species that were transcribed *in vitro*, except that there were no modifications on the transcripts *in vitro*. The pseudouridine (Ψ) bases shown in boxes would be expected to be modified by mPus1p *in vitro*. The other pseudouridine bases in the tRNA species are not modified by mPus1p. The arrows on the $tRNA^{Ile}$ sequence mark the ends of the intron.

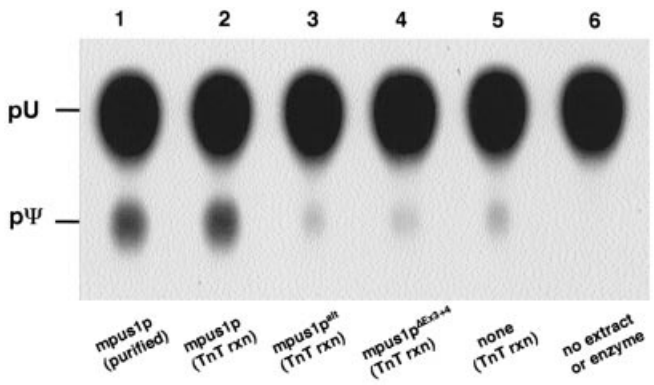


Figure 9 mPus1p, unlike the isoforms $mPus1p^{alt}$ and $mPus1p^{\Delta Ex3+4}$, forms pseudouridine on $tRNA^{Val}$

Autoradiograph of a portion of a TLC plate of ^{32}P-labelled tRNA incubated with purified recombinant mPus1p (lane 1), mPus1p translated *in vitro* (lane 2), $mPus1p^{alt}$ translated *in vitro* (lane 3), $mPus1p^{\Delta Ex3+4}$ translated *in vitro* (lane 4), lysate alone (lane 5) or no enzyme or lysate (lane 6). After incubation, the tRNA was isolated, digested with nuclease P1 and chromatographed on cellulose plates in one dimension (see the Experimental section for details). The positions of uridine 5′-monophosphate (pU), and pseudouridine 5′-monophosphate (pΨ) are indicated at the left.

44 kDa. The approx. 37 kDa band might correspond to a protein translated from an mRNA represented by cDNA C (Figure 1; predicted size 38931 Da). In heart there is a minor band at approx. 52 kDa in addition to a band at 44 kDa. This larger minor band might correspond to a protein product containing both the additional 18 residues of exon IV and the additional 30 residues found in exon I (see the above description of cDNA A). In the sample from liver there was a band at approx. 44 kDa but none of the other predicted isoforms were observed. There were additional bands at more than 100 kDa in all the lanes of the Western blot other than in the heart sample lane. These bands did not correspond to the predicted size of an isoform and were of unknown origin. This was the only protein reactive with the antibody in the kidney sample and it is possible that there is a large protein in mouse that shares an epitope with mPus1p.

Pseudouridine synthase activity of isoforms of mPus1p translated *in vitro*

The ORF for the mPus1p isoform that included the additional 18 residues from exon VI ($mPus1p^{alt}$) and the ORF lacking the coding region from exons III and IV ($mPus1p^{\Delta Ex3+4}$) were inserted into the expression vector pET16b. The resultant recombinant proteins has a leader sequence (21 residues; see the Experimental section) containing a string of histidine residues (see the Experimental section) that does not inhibit the pseudouridine synthase activity of mPus1p [7]. Recombinant mPus1p was expressed in *E. coli* and isolated by Ni^{2+}-nitrilotriacetate column chromatography from cleared supernatants of lysed cells [7], but the two isoforms ($mpus1^{alt}$ and $mPus1p^{\Delta Ex3+4}$) were in inclusion bodies and were found in the pellet after cell lysis. Isolating the proteins from the pellet would require the use of denaturing conditions; the proteins would then have to be renatured before use in the synthase assays. Because of concern that the proteins might not renature properly, they were synthesized *in vitro* with a coupled transcription–translation system instead (see the Experimental section). Aliquots of each reaction mixture were subjected to SDS/PAGE and blotted to nitrocellulose. The blot was probed with the same anti-mPus1p antibody as that used for Figure 6. All three of the mPus1p isoforms reacted with the antibody and were at approximately the same concentration in the reactions (Figure 7). Purified recombinant mPus1p expressed

Table 1 Pseudouridine synthase activity in rabbit reticulocyte extracts expressing mPus1p and mPus1p isoforms

Aliquots of coupled transcription–translation reactions (Promega TnT® Coupled Reticulocyte Lysate system) were added to the substrate RNA species (see the Experimental section) and incubated for 20 min at 37 °C. Plasmid DNA species for the expression of mPus1p or mPus1p isoforms were used in the transcription–translation reaction; however, in the last row no plasmid DNA was added to the lysate. For yeast $tRNA^{Phe}$ the values shown were calculated by scraping the spot off the plate and counting it. The raw c.p.m. were corrected for background (no extract or enzyme). The number of moles of pseudouridine per mole of tRNA was calculated from the c.p.m. in the uridine 5′-monophosphate and pseudouridine 5′-monophosphate spots and the number of total uridines and pseudouridines found in the substrate. Listed alongside the substrate in parentheses is the number of pseudouridines (Ψ) in that substrate that could be modified by mPus1p (see Figure 8). Human $tRNA^{Ile}$ has an intron (see Figure 8).

		Pseudouridine synthase activity (mol of pseudouridine/mol of tRNA)		
Enzyme	Substrate…	Yeast $tRNA^{Phe}$ (0Ψ)	Yeast $tRNA^{Val}$ (1Ψ)	Human $tRNA^{Ile}$ (3Ψ)
mPus1p*		< 0.01	0.79	1.31
mPus1p translated *in vitro*		0.02	0.79	1.33
$mPus1p^{alt}$ translated *in vitro*		0.02	0.13	0.34
$mPus1p^{\Delta Ex3+4}$ translated *in vitro*		0.02	0.16	0.33
No enzyme (lysate alone)		0.02	0.20	0.22

* The purified mPus1p was mixed with an aliquot of lysate alone in these samples.

in *E. coli* (approx. 30 ng; Figure 7, lane 1) was included on the blot to estimate the amount of the protein made for each of the isoforms (approx. 10 ng). The proteins were slightly larger than expected because there was a 21-residue leader on the N-terminus of each protein (see the Experimental section).

Aliquots of the resultant proteins translated *in vitro* were used in pseudouridine synthase assays that used the [^{32}P]UTP-labelled unmodified tRNA substrates shown in Figure 8. Two of these, $tRNA^{Val}$ and $tRNA^{Ile}$, are substrates that mPus1p is known to modify *in vitro*, whereas $tRNA^{Phe}$ is not modified by mPus1p [7]. The substrates were incubated for 20 min with purified recombinant mPus1p, mPus1p translated *in vitro*, $mPus1p^{alt}$ translated *in vitro* or $mPus1p^{\Delta Ex3+4}$ translated *in vitro*. The RNA was isolated from the reaction mixture, digested with nuclease P1 and chromatographed on TLC plates in one dimension. An example of the resultant autoradiograph, for the experiment with $tRNA^{Val}$ as the substrate, is shown in Figure 9. It can be seen that the activity was high when either mPus1p preparation was incubated with $tRNA^{Val}$ but there was little activity in any of the other reaction mixtures. There was a small amount of activity in the reticulocyte extract alone; the two isoforms exhibited no activity above that level. This same procedure was repeated for all the substrates (Figure 8) and the amount of radioactivity was determined by scraping the spots and scintillation counting. The results are compiled in Table 1 and again it is apparent that neither isoform had activity towards the substrates recognized by mPus1p.

DISCUSSION

Evidence in the form of fully sequenced ESTs, ribonuclease protection analysis, and RT–PCR assays is presented for three alternatively spliced forms of the message for mPus1p. It seems that at least three proteins are translated in either mouse tissues or cultured RAW264.7 cells. One protein matches the size of that already published for mPus1p [7]; another, approx. 2 kDa larger, corresponds to the size expected if the 18 residues of exon IV were included in mPus1p. There is also evidence that a protein that would result from the translation of an mRNA that is lacking both exons III and IV is expressed in the spleen of the mouse. In addition, the heart contains a minute amount of an isoform that might correspond to a protein product containing both the additional 18 residues of exon IV and the additional 30 residues found in exon I. The differential expression of the mPus1p isoforms in the tissues of the mouse and in various mouse cell lines will be an avenue of future investigation.

The two isoforms, $mPus1p^{alt}$ and $mPus1p^{\Delta Ex3+4}$, that were expressed *in vitro* did not form pseudouridine when incubated with yeast $tRNA^{Val}$ and $tRNA^{Ile}$, two known substrates of mPus1p [7]. These results are not surprising given that both of these mPus1p isoforms have changes in the primary sequence near the absolutely conserved aspartate residue that is embedded in the RTDKGV sequence, a highly conserved motif in the truA family members [19,38]. This aspartate residue is known to be essential for the activity of PSUI, a truA family member from *E. coli* [38], and in representatives of other pseudouridine synthase families [47,48]. When the crystal structure of PSUI was solved, this aspartate residue was found at the centre of the RNA-binding cleft of the protein [39]. The $mPus1p^{\Delta Ex3+4}$ isoform is lacking the exon containing the aspartate residue and therefore was not expected to be active. The $mPus1p^{alt}$ isoform inserts 18 residues (PLRTSACFCLYTGFLFHW) into the cleft after the lysine residue (K) in the RTDKGV sequence. The insert has a net positive charge and hydrophobic character, similar to that of the cleft in PSUI [39]. However, it is reasonable to assume that this additional sequence near the active site of the enzyme would disrupt at least the recognition of the substrate, if not the catalysis step [39].

Given these results, why would the cells make the $mPus1p^{alt}$ or $mPus1p^{\Delta Ex3+4}$ form? For $mPus1p^{alt}$, it is possible that the true substrate for this enzyme has not been found. The additional amino acids inserted near the active site might change the specificity of the enzyme by changing the distribution of positive charges in the cleft that binds the RNA substrate [39]. It is worth noting that even though there is considerable identity between mPus1p and PSUI (23 %) [7] and there are stretches of near-perfect identity, these two enzymes do not modify the same regions on tRNA substrates. PSUI modifies the 38, 39 or 40 residues of *E. coli* tRNA [10], whereas mPus1p modifies 27, 28, 34 and 36 residues in yeast and mammalian tRNA species [7]. The differences in the primary sequences of these two enzymes contribute to the difference in the specificities and it is possible that $mPus1p^{alt}$ recognizes a substrate that is as yet undetermined.

The alternatively spliced forms might be regulators of mPus1p activity. From the crystal structure it has been shown that the active form of PSUI is probably the dimer [39]. If mPus1p is also a dimer in the active form, a heterodimer between mPus1p and $mPus1p^{alt}$ or $mPus1p^{\Delta Ex3+4}$ might be inactive. From the opposite

perspective it is possible that the activity of mPus1p is modulated when dimerized with another isoform, not necessarily inactivated. It has been shown with yeast Pus1p that oligomerization of the enzyme at high concentrations inhibits activity [49]. It is possible that the presence of mPus1p^{alt} or mPus1p$^{\Delta Ex3+4}$ would inhibit higher-order oligomerization of mPus1p and could actually increase the efficiency of the enzyme. Obviously because we have the ability to produce these proteins *in vitro* we shall be able to test these possibilities.

Finally, one can take the cynical view that mPus1p^{alt} and mPus1p$^{\Delta Ex3+4}$ are just the result of fortuitous mutations in the genome of the mouse. These mutations permit the inclusion of exon IV into, or the deletion of exons III and IV from, RNA in a statistically significant number of transcripts. The protein is then translated from those transcripts and the product is stable in cells, possibly because it forms a dimer with mPus1p. Determining whether exon IV and the splicing pathways seen in the mouse are conserved in other mammals should answer this question.

Almost 3700 bp of sequence upstream of the transcription start site was determined for the *MPUS*1 gene. When the sequence immediately upstream (1 kb) was subjected to analysis for transcription-factor-binding sites (http://www.cbil.upenn.edu/cgi-bin/tess/tess33?RQ = SEA-FR-Query), several points emerged. First, there seems to be no TATA or TATA-like box near the start of transcription, even when the start site is shifted upstream by 45 nt. The sequence around the identified start of transcription for the *MPUS*1 gene is not a good match for the initiator element that has been identified in both TATA-less and TATA-containing promoters [50]. The consensus initiator sequence is PyPyA(+1)NT/APyPy [50] but the MPUS1 sequence is TCG(+1)TGCA. It is possible that the correct 5′ end has not been identified with the use of primer extension, but these experiments are difficult in view of the low abundance and alternative splicing at the 5′ end of the message.

Now that the structure and sequence of the *MPUS*1 gene is known, it will be possible to study the effect of the targeted disruption of this gene in the mouse. Deletion of the *PSUI* gene [23], the *RluD* gene [13] and the *RluA* gene [22] in *E. coli*, and the deletion of the *PUS*3 gene in yeast [12], result in considerably slower growth rates for the deleted strains. When the yeast *PUS*1 gene was deleted there was no appreciable affect on growth [14]. However, mPus1p is the first pseudouridine synthase from metazoans to be cloned and it is difficult to predict the effect of the absence of this modification enzyme on the development of the mouse. The experiment will need to be done and the results should be interesting.

We thank Dr O. Uhlenbeck for the yeast tRNA^{Phe} clone and Dr F. Fasiolo and Dr H. Grosjean for the yeast tRNA^{Val} (with Met anticodon UAC) clone and the tRNA^{Ile} (with intron) clone. This work was supported by grants to J.R.P. from the National Science Foundation (MCB-9603550) and from the Research Development Fund of the University of South Carolina School of Medicine.

REFERENCES

1 Bjork, G. R., Ericson, J. U., Gustafsson, C. E., Hagervall, T. G., Jonsson, Y. H. and Wikstrom, P. M. (1987) Transfer RNA modification. Annu. Rev. Biochem. **56**, 263–287
2 Ofengand, J. and Fournier, M. J. (1998) The pseudouridine residues of rRNA: number, location, biosynthesis, and function. In Modification and Editing of RNA (Grosjean, H. and Benne, R., eds.), pp. 229–253, ASM Press, Washington, DC
3 Massenet, S. A., Mougin, A. and Branlant, C. (1998) Posttranscriptional modifications of the U small nuclear RNAs. In Modification and Editing of RNA (Grosjean, H. and Benne, R., eds.), pp. 201–227, ASM Press, Washington, DC
4 Johnson, L. and Soll, D. (1970) *In vitro* biosynthesis of pseudouridine at the polynucleotide level by an enzyme extract from *Escherichia coli*. Proc. Natl. Acad. Sci. U.S.A. **67**, 943–950
5 Ansmant, I., Massenet, S., Grosjean, H., Motorin, Y. and Branlant, C. (2000) Identification of the *Saccharomyces cerevisiae* RNA:pseudouridine synthase responsible for formation of Ψ2819 in 21S mitochondrial ribosomal RNA. Nucleic Acids Res. **28**, 1941–1946
6 Becker, H. F., Motorin, Y., Planta, R. J. and Grosjean, H. (1997) The yeast gene YNL292w encodes a pseudouridine synthase (Pus4) catalyzing the formation of Ψ55 in both mitochondrial and cytoplasmic tRNAs. Nucleic Acids Res. **25**, 4493–4499
7 Chen, J. and Patton, J. R. (1999) Cloning and characterization of a mammalian pseudouridine synthase. RNA **5**, 409–419
8 Conrad, J., Sun, D., Englund, N. and Ofengand, J. (1998) The rluC gene of *Escherichia coli* codes for a pseudouridine synthase that is solely responsible for synthesis of pseudouridine at positions 955, 2504, and 2580 in 23 S ribosomal RNA. J. Biol. Chem. **273**, 18562–18566
9 Huang, L., Ku, J., Pookanjanatavip, M., Gu, X., Wang, D., Greene, P. J. and Santi, D. V. (1998) Identification of two *Escherichia coli* pseudouridine synthases that show multisite specificity for 23S RNA. Biochemistry **37**, 15951–15957
10 Kammen, H. O., Marvel, C. C., Hardy, L. and Penhoet, E. E. (1988) Purification, structure, and properties of *Escherichia coli* tRNA pseudouridine synthase I. J. Biol. Chem. **263**, 2255–2263
11 Lafontaine, D. L. J., Bousquet-Antonelli, C., Henry, Y., Caizergues-Ferrer, M. and Tollervey, D. (1998) The box H + ACA snoRNAs carry Cbf5p, the putative rRNA pseudouridine synthase. Genes Dev. **12**, 527–537
12 Lecointe, F., Simos, G., Sauer, A., Hurt, E. C., Motorin, Y. and Grosjean, H. (1998) Characterization of yeast protein Deg1 as pseudouridine synthase (Pus3) catalyzing the formation of Ψ38 and Ψ39 in tRNA anticodon loop. J. Biol. Chem. **273**, 1316–1323
13 Raychaudhuri, S., Conrad, J., Hall, B. G. and Ofengand, J. (1998) A pseudouridine synthase required for the formation of two universally conserved pseudouridines in ribosomal RNA is essential for normal growth of *Escherichia coli*. RNA **4**, 1407–1417
14 Simos, G., Tekotte, H., Grosjean, H., Segref, A., Sharma, K., Tollervey, D. and Hurt, E. C. (1996) Nuclear pore proteins are involved in the biogenesis of functional tRNA. EMBO J. **15**, 2270–2284
15 Wrzesinski, J., Nurse, K., Bakin, A., Lane, B. G. and Ofengand, J. (1995) A dual-specificity pseudouridine synthase: an *Escherichia coli* synthase purified and cloned on the basis of its specificity for Ψ746 in 23S RNA is also specific for Ψ32 in tRNA(phe). RNA **1**, 437–448
16 Nurse, K., Wrzesinski, J., Bakin, A., Lane, B. G. and Ofengand, J. (1995) Purification, cloning, and properties of the tRNA Ψ55 synthase from *Escherichia coli*. RNA **1**, 102–112
17 Wrzesinski, J., Bakin, A., Nurse, K., Lane, B. G. and Ofengand, J. (1995) Purification, cloning, and properties of the 16S RNA pseudouridine 516 synthase from *Escherichia coli*. Biochemistry **34**, 8904–8913
18 Niu, L. and Ofengand, J. (1999) Cloning and characterization of the 23S RNA pseudouridine 2633 synthase from *Bacillus subtilis*. Biochemistry **38**, 629–635
19 Koonin, E. V. (1996) Pseudouridine synthases: four families of enzymes containing a putative uridine-binding motif also conserved in dUTPases and dCTP deaminases. Nucleic Acids Res. **24**, 2411–2415
20 Motorin, Y., Keith, G., Simon, C., Foiret, D., Simos, G., Hurt, E. and Grosjean, H. (1998) The yeast tRNA:pseudouridine synthase Pus1p displays a multisite substrate specificity. RNA **4**, 856–869
21 Massenet, S., Motorin, Y., Lafontaine, D. L., Hurt, E. C., Grosjean, H. and Branlant, C. (1999) Pseudouridine mapping in the *Saccharomyces cerevisiae* spliceosomal U small nuclear RNAs (snRNAs) reveals that pseudouridine synthase pus1p exhibits a dual substrate specificity for U2 snRNA and tRNA. Mol. Cell. Biol. **19**, 2142–2154
22 Raychaudhuri, S., Niu, L., Conrad, J., Lane, B. G. and Ofengand, J. (1999) Functional effect of deletion and mutation of the *Escherichia coli* ribosomal RNA and tRNA pseudouridine synthase RluA. J. Biol. Chem. **274**, 18880–18886
23 Tsui, H. C., Arps, P. J., Connolly, D. M. and Winkler, M. E. (1991) Absence of hisT-mediated tRNA pseudouridylation results in a uracil requirement that interferes with *Escherichia coli* K-12 cell division. J. Bacteriol. **173**, 7395–7400
24 Inoue, K., Ohno, M. and Shimura, Y. (1995) Aspects of splice site selection in constitutive and alternative pre-mRNA splicing. Gene Expr. **4**, 177–182
25 Chabot, B. (1996) Directing alternative splicing: cast and scenarios. Trends Genet. **12**, 472–478
26 Lopez, A. J. (1998) Alternative splicing of pre-mRNA: developmental consequences and mechanisms of regulation. Annu. Rev. Genet. **32**, 279–305
27 Lennon, G., Auffray, C., Polymeropoulos, M. and Soares, M. B. (1996) The I.M.A.G.E. Consortium: an integrated molecular analysis of genomes and their expression. Genomics **33**, 151–152
28 Sanger, F., Nicklen, S. and Coulson, A. R. (1977) DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. U.S.A. **74**, 5463–5467

29 Sampson, J. R. and Uhlenbeck, O. C. (1988) Biochemical and physical characterization of an unmodified yeast phenylalanine transfer RNA transcribed *in vitro*. Proc. Natl. Acad. Sci. U.S.A. **85**, 1033–1037
30 Szweykowska-Kulinska, Z., Senger, B., Keith, G., Fasiolo, F. and Grosjean, H. (1994) Intron-dependent formation of pseudouridines in the anticodon of *Saccharomyces cerevisiae* minor tRNA(Ile). EMBO J. **13**, 4636–4644
31 Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd edn, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
32 Mullis, K. B. and Faloona, F. A. (1987) Specific synthesis of DNA *in vitro* via a polymerase-catalyzed chain reaction. Methods Enzymol. **155**, 335–350
33 Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T_4. Nature (London). **227**, 680–685
34 Melton, D. A., Krieg, P. A., Rebagliati, M. R., Maniatis, T., Zinn, K. and Green, M. R. (1984) Efficient in vitro synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. Nucleic Acids Res. **12**, 7035–7056
35 Chomczynski, P. and Sacchi, N. (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal. Biochem. **162**, 156–159
36 Church, D. M. and Buckler, A. J. (1999) Gene identification by exon amplification. Methods Enzymol. **303**, 83–99
37 Nishimura, S. (1972) Minor components in transfer RNA: their characterization, location, and function. Prog. Nucleic Acid Res. Mol. Biol. **12**, 49–85
38 Huang, L., Pookanjanatavip, M., Gu, X. and Santi, D. V. (1998) A conserved aspartate of tRNA pseudouridine synthase is essential for activity and a probable nucleophilic catalyst. Biochemistry **37**, 344–351
39 Foster, P. G., Huang, L., Santi, D. V. and Stroud, R. M. (2000) The structural basis for tRNA recognition and pseudouridine formation by pseudouridine synthase I. Nat. Struct. Biol. **7**, 23–27

Received 31 May 2000/11 September 2000; accepted 29 September 2000

40 Wahle, E. and Ruegsegger, U. (1999) 3′-End processing of pre-mRNA in eukaryotes. FEMS Microbiol. Rev. **23**, 277–295
41 Robberson, B. L., Cote, G. J. and Berget, S. M. (1990) Exon definition may facilitate splice site selection in RNAs with multiple exons. Mol. Cell. Biol. **10**, 84–94
42 Berget, S. M. (1995) Exon recognition in vertebrate splicing. J. Biol. Chem. **270**, 2411–2414
43 Sterner, D. A., Carlo, T. and Berget, S. M. (1996) Architectural limits on split genes. Proc. Natl. Acad. Sci. U.S.A. **93**, 15081–15085
44 Hodgson, C. P., Fisk, R. Z., Arora, P. and Chotani, M. (1990) Nucleotide sequence of mouse virus-like (VL30) retrotransposon BVL-1. Nucleic Acids Res. **18**, 673
45 Berger, J., Truppe, C., Neumann, H. and Forss-Petter, S. (1998) cDNA cloning and mRNA distribution of a mouse very long-chain acyl-CoA synthetase. FEBS Lett. **425**, 305–309
46 Hirsch, D., Stahl, A. and Lodish, H. F. (1998) A family of fatty acid transporters conserved from mycobacterium to man. Proc. Natl. Acad. Sci. U.S.A. **95**, 8625–8629
47 Ramamurthy, V., Swann, S. L., Paulson, J. L., Spedaliere, C. J. and Mueller, E. G. (1999) Critical aspartic acid residues in pseudouridine synthases. J. Biol. Chem. **274**, 22225–22230
48 Conrad, J., Niu, L., Rudd, K., Lane, B. G. and Ofengand, J. (1999) 16S ribosomal RNA pseudouridine synthase RsuA of *Escherichia coli*: deletion, mutation of the conserved Asp102 residue, and sequence comparison among all other pseudouridine synthases. RNA **5**, 751–763
49 Arluison, V., Batelier, G., Ries-Kautt, M. and Grosjean, H. (1999) RNA:pseudouridine synthetase Pus1 from *Saccharomyces cerevisiae*: oligomerization property and stoichiometry of the complex with yeast tRNA(Phe). Biochimie **81**, 751–756
50 Smale, S. T. (1997) Transcription initiation from TATA-less promoters within eukaryotic protein-coding genes. Biochim. Biophys. Acta **1351**, 73–88