# Mitochondrial aspartate aminotransferase: direction of a single protein with two distinct functions to two subcellular sites does not require alternative splicing of the mRNA

Michael W. BRADBURY[1] and Paul D. BERK

Department of Medicine (Division of Liver Diseases), Mount Sinai School of Medicine, Box 1633, 1 Gustave L. Levy Place, New York, NY 10029, U.S.A.

During differentiation of mouse 3T3-L1 fibroblasts to an adipocyte phenotype, the mitochondrial isoform of aspartate aminotransferase accumulates on the plasma membrane. The determination of whether this reflects translation of an alternatively spliced message lacking the mitochondrial leader sequence required cloning of the enzyme's uncommon *a* allele, for which these cells are homozygous. The 1.4-kb cDNA sequence of the *a* allele was obtained from oligo-dT-primed reverse-transcriptase PCR products amplified from FVB mouse RNA. It differed from the *b* allele at only 2 bp and one amino acid. By contrast, gene-specific primers generated an additional 1.4-kb fragment that differed from the *b* allele by $\approx 1\,\%$ of nucleotides, encoding four amino acid substitutions. This sequence proved to represent a recently diverged processed pseudogene. The presence of such pseudogenes can complicate interpretation of expressed-sequence-tag data and single-nucleotide-polymorphism genotyping studies. Using probes derived from the *a* allele, RNase protection analyses indicated that only a single message for the enzyme was present in 3T3-L1 fibroblasts and adipocytes, despite differences in subcellular protein distribution.

Key words: cDNA, plasma-membrane fatty acid-binding protein, polymorphism, pseudogene, signal peptide.

## INTRODUCTION

Mitochondrial aspartate aminotransferase (mAspAT; EC 2.6.1.1), a common enzyme found in the mitochondria of virtually all cells, plays an important role in amino acid metabolism and is a critical component of the malate-aspartate shuttle, which transports reducing equivalents across the inner mitochondrial membrane. Recently, it has also been identified by immunohistological, surface labelling and immunoprecipitation techniques on the plasma membranes of hepatocytes, adipocytes and other cells with high transmembrane fluxes of long-chain free fatty acids [1–7]. There it has been proven identical to the protein designated plasma-membrane fatty acid-binding protein (FABPpm) [8,9], which facilitates cellular free fatty acid uptake [1,2,10–13].

Expression of mAspAT mRNA and mAspAT protein synthesis increase dramatically in 3T3-L1 preadipocytes during differentiation from a fibroblast to an adipocyte phenotype. During this process mAspAT, which is barely detectable on the surface of fibroblasts, becomes abundantly expressed on the plasma membrane, and cellular free fatty acid uptake increases in parallel [10,11]. The 3T3-L1 cell line proved to be homozygous for the *a* allele of *mAspAT*, encoded by the *Got-2* (glutamate-oxaloacetate transaminase) locus, whereas all previously cloned sequences for the cDNA are apparently derived from the common *b* allele [14,15]. Therefore, to facilitate studies of mAspAT mRNA expression during 3T3-L1 differentiation, we cloned a cDNA for the *a* allele to provide reagents for procedures such as RNase protection assays (RPAs). During these studies we discovered a recently evolved, previously unrecognized *mAspAT* pseudogene.

A mitochondrial leader sequence is present at the N-terminus of the protein. This leader sequence is coded for by exon 1 of the gene, and the first codon for the mature peptide is interrupted by intron 1 [15]. One possible mechanism for sorting the protein to different compartments would be for alternative splicing to produce a transcript with a different first exon, coding for a different type of leader sequence, thus directing the protein to a different location. Our data indicate that this is not the case, and that the increase in mAspAT on the surface is derived from protein that originally possessed a mitochondrial leader sequence.

## EXPERIMENTAL

### Tissues and reagents

Liver and kidney tissue from FVB and C57BL/6 mice were obtained from Dr. Jon Gordon (Mount Sinai School of Medicine, New York, U.S.A.). Oligonucleotides were prepared by Genset (La Jolla, CA, U.S.A.). Restriction endonucleases and DNA-modifying enzymes were purchased from New England Biolabs (Beverly, MA, U.S.A.). PCR, reverse-transcriptase (RT)-PCR, and cycle-sequencing kits were purchased from PE Biosystems (Norwalk, CT, U.S.A.). Radioisotopes were purchased from NEN Life Science Products (Boston, MA, U.S.A.).

### Cloning of products from F33/R1440-primed RT-PCR

Nucleic acids were prepared from liver tissue of FVB (*a*/*a*) and C57BL/6 (*b*/*b*) mice with Tri-Reagent (Molecular Research Center, Cincinnati, OH, U.S.A.) according to the manufacturer's instructions. Primers used to amplify the nucleic acids encoding the mouse *Got-2* locus were (5′ → 3′): F33 (ACCATCCACTGCCGTCTTAC), representing bases 33–52 of the *b* allele 5′-untranslated region (UTR); and R1440 (ACCACCCACCCTC-

Abbreviations used: EST, expressed sequence tag; *Got-2*, glutamate-oxaloacetate transaminase; mAspAT, mitochondrial aspartate aminotransferase; pmAspAT, pre-mAspAT; NIRCA, non-isotopic RNase cleavage assay; RPA, RNase protection assay; RT, reverse transcriptase; SNP, single nucleotide polymorphism; UTR, untranslated region.

[1] To whom correspondence should be addressed (e-mail michael-bradbury@smtplink.mssm.edu).

CCTCACCCTGC), complementary to bases 1440–1415 of the corresponding 3′-UTR. Liver RNA (5 $\mu$g) from each strain was employed in RT-PCR reactions (RNA PCR Kit, PE Biosystems) using R1440 to prime the RT reaction. Nucleotides, RT, RNAse inhibitor and mineral oil were added according to the manufacturer's directions. Samples were incubated in a Perkin Elmer Thermal Cycler for 15 min at 42 °C, followed by a 5-min denaturation step at 99 °C. Tubes were placed on ice and 80 $\mu$l of PCR reaction mixture was added, composed of buffer, $MgCl_2$, 2.5 units of *Taq* DNA polymerase and 100 ng of each primer. After a 2-min denaturation at 95 °C, the samples were amplified for 35 cycles consisting of a 1-min 95 °C denaturation step followed by a 1-min 60 °C annealing and synthesis step augmented by a 72 °C synthesis step starting at 5 s and increasing by 3 s each cycle, concluding with a final 7 min at 72 °C. Products were analysed by gel electrophoresis. cDNAs were ligated to either pGEM-T vector (Promega, Madison, WI, U.S.A.) or *Sma*I-digested pBluescript II SK(+) (Stratagene, La Jolla, CA, U.S.A.), which were used to transform TOP10 cells. Two clones with the 1.4-kb insert (pRT1, pRT14), one in each vector, were obtained from the FVB sample. The inserts were sequenced with the AmpliCycle Kit (PE Biosystems) and [$\alpha$-$^{33}$P]dATP (NEN Life Science Products). Films were scanned on a *pdi* Scanning Densitometer (*pdi*, Huntington Station, NY, U.S.A.) attached to a Sun Sparc Station, and analysed using DNA Code software (*pdi*).

### Direct sequencing and cloning of oligo-dT-primed RT-PCR product

RT-PCR products were generated as described above from FVB liver RNA, with oligo-dT replacing R1440 to prime first-strand synthesis, and separated on 1.2% agarose or 45% Oligo-Prep gels (National Diagnostics, Atlanta, GA, U.S.A.). The 1.4-kb product was isolated from the gels using the GeneClean kit (Bio 101, Vista, CA, U.S.A.) for Oligo-Prep gels or the Qiagen Gel Extraction kit (Qiagen, Valencia, CA, U.S.A.) for agarose gels. The product was sequenced at the Utah State University Biotechnology Center (Logan, UT, U.S.A.). Gel-purified insert was also cloned into the pGEM-T Easy vector (Promega). A series of clones with 1.4-kb inserts were produced, and several were sequenced partially by the Utah State University Biotechnology Center to confirm their identities as the cDNA for the *a* allele.

### Testing for DNA contamination of RT-PCR template

RT-PCR reactions were carried out with the following modifications. RT was omitted to prevent first-strand cDNA synthesis, which should result in no product. RNAse-free DNAse I (Promega) was added to some samples in place of RT to digest any contaminating DNA. In these tubes only RNA would be present at the initiation of PCR.

### Genomic PCR

Genomic DNA (1 $\mu$g) from FVB or C57BL/6 liver was amplified in a 100-$\mu$l reaction mixture with 100 ng of primer F33, 120 ng of primer R1440, buffer, $MgCl_2$, deoxynucleotides and *Taq* polymerase, as provided in the RNA PCR kit (PE Bisosytems). Cycle parameters were similar to those used in RT-PCR.

### Sequence-data analysis

The sequences obtained were compared with those in GenBank using GCG software (Genetics Computer Group, Madison, WI, U.S.A.) on the Mount Sinai VAX. Protein comparisons were done using BLAST [16] at the National Center for Biotechnology Information (Bethesda, MD, U.S.A.), using the BLAST Network service.

### Non-isotopic RNase cleavage assay (NIRCA) test of identity

A NIRCA assay kit (Mismatch Detect II, Ambion, Austin, TX, U.S.A.) was employed according to the manufacturer's instructions to identify sequence differences between RT-PCR products and plasmid clone pRT14. RT-PCR product was produced from FVB and C57BL/6 liver RNA as described using oligo-dT to prime the first-strand synthesis. RT-PCR products, plasmid pRT14 DNA and genomic DNA from FVB mice were amplified using primers containing T7 (sense) or SP6 (anti-sense) promoter sequences. Primer pairs (all 5′ → 3′) ran from bases 63 to 901 (TF63, TAATACGACTCACTATAGGATGGCTGCTGCCT; SR901, ATTTAGGTGACACTATAGAAGGCTCCCACACG) or bases 705 to 1324 (TF705, TAATACGACTCACTATAGGAGATAGCGTCCG; SR1324, ATTTAGGTGACACTATAGAATTACTTGGTGACCTG). Each of the PCR reaction products was used as a template for RNA production using T7 and SP6 polymerases in separate reactions, according to the manufacturer's directions. All combinations of sense and anti-sense products were hybridized, and digested with two different RNAse mixtures as supplied with the kit and diluted according to the manufacturer's instructions. The first was a mixture of 1/300 RNAse #1 and 1/100 RNAse #2, and the second was 1/100 RNAse #3. After digestion, the reactions were separated on a 2% agarose gel and photographed.

### RPA to detect alternative splicing

A subclone containing a 271-bp region of the *a* allele in pGEM-T Easy, running from the 5′ end of the RT-PCR product to the *Bsp*EI site at the end of exon 2, was prepared in pBluescript II SK(+). This plasmid was linearized with *Not*I, and probe prepared with the MAXIscript T7 kit (Ambion) and [$\alpha$-$^{32}$P]UTP (NEN Life Science Products). This probe was hybridized to total RNA from 3T3-L1 fibroblasts and 4- or 8-day differentiating adipocytes prepared with the RNeasy Mini kit (Qiagen) using the RPA III kit from Ambion. Hybridizations were carried out at temperatures ranging from 37 to 65 °C overnight, and the products were treated with RNAse, precipitated and analysed by denaturing PAGE according to the manufacturer's directions.

## RESULTS

Initial efforts to clone the *a* allele cDNA from a commercial 3T3 library (Clontech, Palo Alto, CA, U.S.A.) by conventional techniques yielded inconsistent results in that overlapping clones did not always contain the same sequence. Electrophoretic analysis [17] demonstrated that, whereas clonally derived 3T3-L1 cells are homozygous for *Got-2 a/a*, 3T3 cells are a heterogeneous mixture of *Got-2 a/a* and *Got-2 a/b* cells (results not shown). Therefore, an RT-PCR strategy was attempted.

Liver RNA from both FVB (*a/a*) and C57BL/6 (*b/b*) mice was amplified by RT-PCR, using oligo-dT primers to selectively amplify mRNA, and the resulting 1.4-kb products were purified. The sequence obtained from the FVB sample, reflecting the cDNA for the *a* allele, differed from the published *b* allele sequence at only two positions (Figure 1). One of these was a silent single nucleotide (C → T) change at the third position of codon 192, which was detected previously in the initial library screening. The other, a C → G change at base position 496, resulted in a Q146E mutation. The change from a glutamine to a glutamate could account for the defining electrophoretic mobility shift of the *a* allele [17].

```
---ACCATCCACTGCCGTCTTACCGCCCACC ATG GCC CTC CTG CAC TCC AGC CGC ATC CTC TCC GGG ATG GCT GCT GCC TTT CAC CCA GGC CTA GCT   94
                               Met Ala Leu Leu His Ser Ser Arg Ile Leu Ser Gly Met Ala Ala Ala Phe His Pro Gly Leu Ala   22

                                                          C
GCT GCA GCC TCT GCC AGA GCC AGC TCC TGG TGG ACC CAT GTT GAA ATG GGA CCT CCA GAT CCC ATC CTG GGC GTT ACC GAA GCC TTC AAG  184
Ala Ala Ala Ser Ala Arg Ala Ser Ser Trp Trp Thr His Val Glu Met Gly Pro Pro Aps Pro Ile Leu Gly Val Thr Glu Ala Phe Lys   52

                                                                                                          T
AGA GAT ACC AAC AGC AAG AAG ATG AAC CTG GGA GTT GGT GCC TAC CGG GAT GAT AAC GGA AAA CCT TAC GTG CTC CCC AGT GTC CGG AAG  274
Arg Asp Thr Asn Ser Lys Lys Met Asn Leu Gly Val Gly Ala Tyr Arg Asp Asp Asn Gly Lys Pro Tyr Val Leu Pro Ser Val Arg Lys   82

                                                 G                                                                A
GCA GAG GCC CAG ATT GCT GCA AAA AAT TTG GAC AAA GAA TAC CTG CCC ATT GGG GGA CTG GCT GAA TTC TGT AAG GCT TCT GCA GAA CTG  364
Ala Glu Ala Gln Ile Ala Ala Lys Asn Leu Asp Lys Glu Tyr Leu Pro Ile Gly Gly Leu Ala Glu Phe Cys Lys Ala Ser Ala Glu Leu  112

                     C
GCC CTG GGC GAG AAC AAT GAA GTG TTG AAA AGC GGC CGG TTC GTC ACT GTG CAG ACC ATT TCC GGG ACT GGA GCC TTA AGG GTC GGG GCC  454
Ala Leu Gly Glu Asn Asn Glu Val Leu Lys Ser Gly Arg Phe Val The Val Gln The Ile Ser Gly The Gly Ala Leu Arg Val Gly Ala  142

            C                                A                                   G           G
AGT TTT CTG CAA AGG TTT TTT AAG TTC AGC CGA GAT GTC TTT CTG CCC AAA CCA TCC TGG GGA AAT CAC ACG CCC ATC TTC AGG GAT GCC  544
Ser Phe Leu Glu Arg Phe Phe Lys Phe Ser Arg Asp Val Phe Leu Pro Lys Pro Ser Trp Gly Asn His Thr Pro Ile Phe Arg Asp Ala  172

                                                                                        C
GGC ATG CAG CTC CAA GGT TAT CGC TAC TAT GAC CCC AAG ACT TGC GGT TTT GAC TTC TCT GGA GCC CTA GAA GAC ATA TCA AAA ATC CCA  634
Gly Met Gln Leu Gln Gly Tyr Arg Tyr Tyr Asp Pro Lys Thr Cys Gly Phe Asp Phe Ser Gly Ala Leu Glu Asp Ile Ser Lys Ile Pro  202

                                 A                                 A                 T
GAG CAG AGT GTC CTC CTC CTG CAT GCC TGC GCT CAC AAC CCC ACC GGC GTG GAC CCG CGT CCC GAG CAG TGG AAG GAG ATA GCG TCC GTG  724
Glu Gln Ser Val Leu Leu Leu His Ala Cys Als His Asn Pro Thr Gly Val Asp Pro Arg Pro Glu Gln Trp Lys Glu Ile Ala Ser Val  232

GTG AAG AAA AAG AAT CTC TTC GCA TTC TTT GAC ATG GCC TAC CAA GGC TTT GCC AGC GGT GAT GGT GAT AAG GAT GCC TGG GCC GTG CGG  814
Val Lys Lys Lys Asn Leu Phe Ala Phe Phe Asp Met Ala Tyr Gln Gly Phe Ala Ser Gly Asp Gly Asp Lys Asp Ala Trp Ala Val Arg  262

CAC TTC ATC GAG CAG GGC ATC AAT GTC TGC CTC TGC CAA TCG TAT GCC AAG AAC ATG GGC CTG TAC GGT GAG CGT GTG GGA GCC TTC ACG  904
His Phe Ile Glu Gln Gly Ile Asn Val Cys Leu Cys Gln Ser Tyr Ala Lys Asn Met Gly Leu Tyr Gly Glu Arg Val Gly Ala Phe Thr  292

GTG GTC TGC AAA GAT GCA GAA GAA GCC AAA AGG GTG GAG TCA CAG CTG AAG ATC TTG ATC CGT CCC CTG TAT TCC AAC CCA CCT CTC AAT  994
Val Val Cys Lys Asp Ala Glu Glu Ala Lys Arg Val Glu Ser Gln Leu Lys Ile Leu Ile Arg Pro Leu Tyr Ser Asn Pro Pro Leu Asn  322

GGG GCC CGG ATC GCA GCA ACC ATC CTG ACT TCT CCA GAC TTG CGG AAG CAA TGG TTG CAA GAG GTG AAA GGC ATG GCT GAC CGC ATC ATC 1084
Gly Ala Arg Ile Ala Ala Thr Ile Leu Thr Ser Pro Asp Leu Arg Lys Gln Trp Leu Gln Glu Val Lys Gly Met Ala Asp Arg Ile Ile  352

AGC ATG AGG ACC CAG CTG GTC TCC AAC CTG AAG AAA GAG GGC TCT TCC CAC AAC TGG CAG CAC ATC ACC GAC CAG ATC GGC ATG TTC TGT 1174
Ser Met Arg Thr Gln Leu Val Ser Asn Leu Lys Lys Glu Gly Ser Ser His Asn Trp Gln His Ile Thr Asp Gln Ile Gly Met Phe Cys  382

                                 G
TTC ACC GGC CTA AAG CCA GAG CAG GTA GAG CGG CTG ACC AAG GAG TTC TCG GTC TAC ATG ACA AAG GAT GGC CGA ATC TCC GTG GCA GGG 1264
Phe Thr Gly Leu lys Pro Glu Gln Val Glu Arg Leu Thr Lys Glu Phe Ser Val Tyr Met Thr Lys Asp Gly Arg Ile Thr Val Ala Gly  412

                                                                                        G
GTC ACC TCT GGC AAT GTG GGC TAC CTT GCC CAT GCC ATT CAC CAG GTC ACC AAG TAA TTCCCAGGTGCAAAGGAACAGAGACCACTTTCCCGACAGCCTT 1364
Val Thr Ser Gly Asn Val Gly Tyr Leu Ala His Ala Ile His Gln Val Thr Lys  *                                                430

                  T
TGCGCTCGTGAGCGTCACGTGCAGGGTGAGGGAGGGTGGGTGGT                                                                      1408
```

**Figure 1 Sequence of the cDNA for the *a* allele of mouse *mAspAT* and the processed pseudogene**

The sequence presented is that of the *a* allele. Underlined regions at the beginning and end define the F33 and R1440 primers respectively. The nucleotides and amino acid in white type reversed out on black represent the differences from the *b* allele sequence, in which the highlighted glutamate is replaced by glutamine, a charge difference consistent with the electrophoretic mobility difference between the isoenzymes. Letters above the sequence represent substitutions in the pseudogene sequence. Nucleotides and amino acids are numbered on the right. The pseudogene is the same as the *b* allele at those points where the *a* and *b* alleles differ, i.e. it contains a C at both positions 464 and 604. Were the pseudogene expressed, the predicted amino acid substitutions from the *b* allele would replace two alanine residues with threonine and two arginines with tryptophan and glutamine. These sequences have been deposited in GenBank under accession numbers MMU82470 for the *a* allele and AF038899 for the pseudogene.

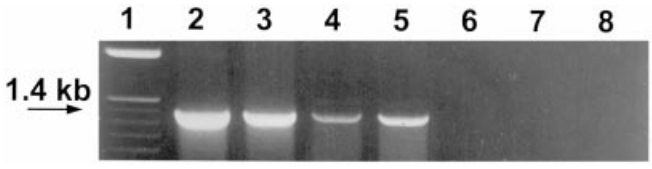


**Figure 2 RT-PCR of mouse total liver RNA**

Lane 1, 100-bp markers. Lanes 2 and 3 are standard RT-PCR samples of FVB (lane 2) and C57BL/6 (lane 3) RNA. Lanes 4 and 5 are samples in which RT was omitted and no cDNA synthesis should have been possible. The presence of amplified product in these lanes suggests that amplification of a pseudogene, possibly from genomic DNA contamination, occurred. Lanes 6 and 7 are samples in which RT was replaced with DNAse. No cDNA synthesis should have occurred to support subsequent PCR, and initial DNA contaminants should have been removed. The lack of amplification product is consistent with the lack of either template. Lane 8, no-RNA control, showing that the source of any template is the RNA preparation, not the reaction mixture.

Total liver RNA from both FVB (*a*/*a*) and C57BL/6 (*b*/*b*) mice also yielded 1.4-kb products when amplified by RT-PCR, with R1440 used to prime first-strand synthesis. Two plasmid clones of the full-length product from FVB RNA (pRT1 and pRT14) were sequenced. The sequences were identical, and differed from the *b* allele by 1% in both RNA and predicted protein sequences, with 14 nucleotide and four amino acid substitutions (Figure 1). However, when these RNA samples were subjected to control reactions in which the RT was replaced with either water or DNase, the results were not as expected for an RNA template. Reduced amplification was found in the case of the samples without RT, no amplification was seen in controls pre-treated with DNase, and no product was seen in the water controls lacking the initial RNA (Figure 2). These results indicate that the plasmid clones could have been generated by amplification of a DNA sequence contaminating the RNA samples. The size and sequence suggested that the source might be a pseudogene in genomic DNA. PCR amplification of genomic DNA of both

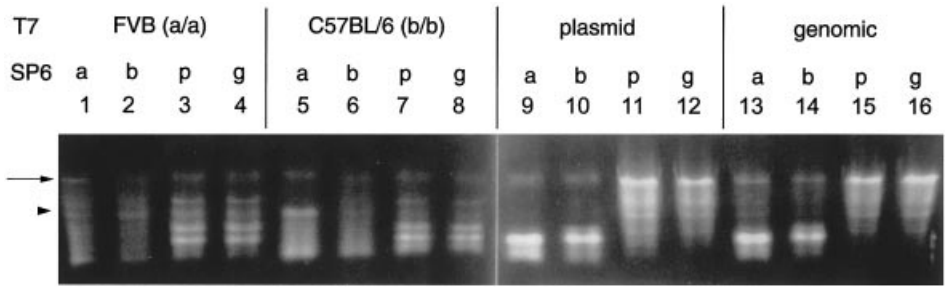


**Figure 3 NIRCA results**

PCR and RT-PCR products from different sources were all amplified using primers containing T7 (sense) or SP6 (anti-sense) promoter sequences. This gel was loaded with samples from the 5′ end (63–901) and digested with RNAse #3 (Mismatch Detect II kit, Ambion). Lanes 1–4 contain FVB sense strand, lanes 5–8 C57BL/6 sense strand, lanes 9–12 sense strand from plasmid pRT14 and lanes 13–16 sense strand from the pseudogene (genomic PCR). Each set of four lanes was hybridized, in order, to FVB (a), C57BL/6 (b), plasmid (p) and pseudogene (g) antisense strands. Lanes 5 and 6 are noteworthy in that they display good protection when hybridizing the two alleles (lane 5) with a prominent central band (arrowhead), and protection of the complete sequence (arrow) in the self-hybridization (lane 6). Lanes 7 and 8 demonstrate that neither the pseudogene nor plasmid sequences are well protected by RNA derived from the *b* allele. However, perfect protection appears to occur when plasmid and pseudogene sequences are hybridized (lanes 11, 12, 15 and 16), arguing for identity of these sequences.

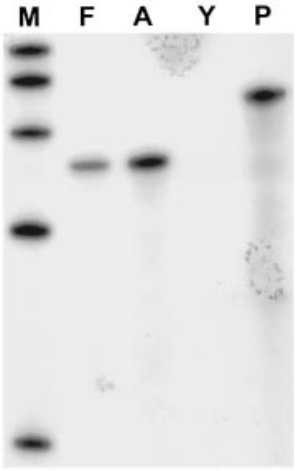


**Figure 4 RPA of 3T3-L1 fibroblast and adipocyte RNA**

Both fibroblast (F) and 4-day differentiating adipocyte (A) RNA samples show a single protected fragment. No protection occurs in hybridizations with yeast RNA (Y). The intact probe (P) is longer than the protected fragment. The protected 271-bp fragment is seen only in the 3T3-L1 samples. There is no evidence of a 153-bp fragment in the adipocyte sample, as would be expected from an mRNA with a different first exon. An increase in the amount of the protected fragment reflects the increase in the amount of mRNA in the 4-day adipocytes. Hybridizations consisted of 10 $\mu$g of total RNA and $10^5$ c.p.m. of probe, and were carried out for 16 h at 47 °C. Molecular-mass markers (M) are $^{32}$P-labelled Century Markers (Ambion) ranging from 100 to 500 bp in length. The autoradiogram was scanned with a Hewlett-Packard 6100CSE scanner into Adobe Photoshop 5.0 for Windows, and extraneous lanes cropped from the Figure.

strains with the same primers also yielded a 1.4-kb product. As the *mAspAT* gene extends over $\approx$ 25 kb with 10 exons and 9 introns, the first intron alone being over 10 kb in length [15], the normal gene could not be the source of this PCR product, again suggesting the existence of a processed pseudogene. In NIRCA analyses, hybridizations of transcripts representing the *a* and *b* alleles exhibited only minor differences, consistent with the sequences obtained. However, when hybridized with one another, the genomic and plasmid sequences showed perfect protection, arguing for identity and confirming that the plasmid clones were representative of the processed pseudogene (Figure 3).

RPA of mAspAT mRNAs from 3T3-L1 fibroblasts and adipocytes (Figure 4) showed perfect protection of a fragment running from the 5′ UTR through to almost the end of exon 2, and no other fragments. Thus there was no evidence that the markedly increased quantities of mAspAT observed on the plasma membranes of differentiated 3T3-L1 adipocytes reflected translation from an alternative message with a different first exon and signal peptide. Quantitative differences were seen reflecting the change in amount of message and protein in differentiating adipocytes [10,11].

## DISCUSSION

mAspAT destined for mitochondria is translated as pre-mAspAT (pmAspAT), with an N-terminal leader that is absent in the mature protein at both its mitochondrial and plasma-membrane sites. Since this leader is encoded by exon 1 of the *mAspAT* gene [15] and the splice site interrupts the first codon of the mature peptide, we hypothesized that alternative splicing of a different first exon might account for sorting of some of the protein to the cell surface. The alternative mRNA, if it exists, should be especially prominent in 3T3-L1 adipocytes, in which cell-surface mAspAT is markedly increased during differentiation from the fibroblast to the adipocyte phenotype [10,11]. If 3T3-L1 mRNA was analysed by RNAse protection, using a probe that spanned the coding sequences for both the leader and the 5′ end of the mature protein, specific results would be expected. The predicted alternative first exon would cause a shorter fragment to be protected in adipocyte mRNA than that expected for the mitochondrial form, which would predominate in fibroblast mRNA. The essential caveat is that the probe used should match perfectly with the fibroblast coding sequence over the selected region. Our finding that 3T3-L1 cells were homozygous for the *a* allele necessitated that the probe be derived from that allele. Since its sequence was unknown, we undertook to clone a cDNA for the *a* allele as a source for the probe. Using such a probe in RPA studies, no shortened protected fragment was found in adipocyte mRNA (Figure 4), indicating that both the mitochondrial and plasma-membrane mAspAT pools are translated from a single pmAspAT message. Because of the 5′ extent of the probe employed, this study does not rule out the existence of a hypothetical additional signal sequence 5′ to the conventional one. However, both cDNA transfection [12] and metabolic labelling experiments [18] argue strongly against this possibility. Further, the 5′ UTR of the *b* allele has no in-frame ATG closer than 471 bp upstream of the known start site, and would code for

only five amino acids before reaching a stop codon if transcription started there. How pmAspAT reaches the plasma membrane as mAspAT remains unclear, but the route may resemble the trans-mitochondrial pathway proposed recently for heat-shock protein 60 (hsp60) [19].

The mouse *mAspAT a* and *b* alleles differ by only 2 bp and one amino acid in the coding region. The Q146E substitution would produce minor changes in secondary structure, and the predicted difference in pI would account for the electrophoretic mobility difference between isoenzymes. Similar one-amino acid differences occur between allelic forms of many human proteins (e.g. [20,21]), although greater divergence has been reported for others [22,23]. Initial efforts to use RT-PCR to amplify the *a* allele from a preparation of total RNA using the F33 and R1440 primers led to the plasmid cloning of a fragment of the appropriate size, with an open reading frame that differed from the *b* allele by $\approx 1\%$ (14/1408 nucleotides and 4/430 amino acids). Rather than an allelic variant, control reactions indicated that this fragment arose from DNA contaminating the initial RNA preparation. The cloned fragment proved identical to a processed pseudogene identified in genomic DNA. That it differs from the *b* allele by less than 1%, a degree of divergence less than that between mouse and rat (6% nucleotide- and 3% protein-sequence divergence), implies that the pseudogene arose after divergence of these species. It also reminds us that such recently diverged pseudogenes may be mistaken for functional alleles under some circumstances.

This study has implications for other current molecular biological methods. The use of expressed sequence tags (ESTs) and libraries is growing [24–26]. As ESTs are derived from cDNA, and our efforts to clone a cDNA led to the cloning of a pseudogene, some ESTs in databases might actually be parts of pseudogenes rather than true cDNA clones. While a search of GenBank EST databases found that all sequences resembling mouse *mAspAT* are more closely related to the two allelic variants than to the pseudogene, this does not ensure that pseudogenes for other genes, or in other species, have not been mistakenly cloned as cDNAs. With respect to single nucleotide polymorphisms (SNPs) [27], there are differences at only two nucleotide positions between the *a* and *b* mouse *mAspAT* alleles, only one of which, at position 464, alters an amino acid. The other is at position 604, and at both positions the *b* allele and the pseudogene are identical. The presence of such a pseudogene could cause difficulties with SNP-mapping protocols. Indeed, in a large study of SNPs [27], many loci were found unsuitable for mapping as the assays did not give the results expected from prior knowledge of sequence polymorphisms and population data. Some of this difficulty may be due to the presence of unrecognized pseudogenes.

In our studies, the cloning of the *a* allele cDNA was necessary for studies of gene expression. The expression studies indicated that both mitochondrial and cell-surface forms of mAspAT were likely to have been derived from a single message pool, with no evidence for an alternatively spliced mRNA. The cloning of the pseudogene was serendipitous, demonstrating the complexity of the genome and that pseudogenes might arise at any point in evolution. Not all of them will have accumulated sufficient mutations to render them obviously non-coding, as this pseudogene retains an open reading frame.

In summary, we have cloned and sequenced the cDNA for the *a* allele of murine *mAspAT*. In addition, we found a recently diverged pseudogene. The cloned cDNA allowed us to examine the mRNA for this gene in 3T3-L1 cells as they underwent differentiation from fibroblasts to adipocytes. The increase in the cell-surface form of the protein results from an increase in mRNA, but there was no evidence for alternative splicing, which was one possible method that cells could use to direct a protein to a different subcellular location. As there do not appear to be two forms of mRNA, there would not be two forms of protein with different signal peptides. These results suggest that, as with hsp60 [19], the surface form of mAspAT may be transported first to the mitochondrion, have the signal peptide removed, and later be transported from the mitochondrion to the cell surface. Further studies addressing this issue are underway.

This work was supported by grants DK-26438 and DK-52401 from the National Institute of Diabetes, Digestive and Kidney Diseases (National Institutes of Health) and by the Mount Sinai Liver Diseases Research Fund.

## REFERENCES


1 Stremmel, W., Strohmeyer, G., Borchard, F., Kochwa, S. and Berk, P. D. (1985) Proc. Natl. Acad. Sci. U.S.A. **82**, 4–8
2 Stremmel, W. and Berk, P. D. (1986) Proc. Natl. Acad. Sci. U.S.A. **83**, 3086–3090
3 Stremmel, W., Strohmeyer, G. and Berk, P. D. (1986) Proc. Natl. Acad. Sci. U.S.A. **83**, 3584–3588
4 Schwieterman, W., Sorrentino, D., Potter, B. J., Rand, J., Kiang, C.-L., Stump, D. D. and Berk, P. D. (1988) Proc. Natl. Acad. Sci. U.S.A. **85**, 359–363
5 Sorrentino, D., Stump, D., Potter, B. J., Robinson, R., White, R., Kiang, C.-L. and Berk, P. D. (1988) J. Clin. Invest. **82**, 928–935
6 Stremmel, W., Lotz, G., Strohmeyer, G. and Berk, P. D. (1985) J. Clin. Invest. **75**, 1068–1076
7 Zhou, S.-L., Gordon, R. E., Bradbury, M., Stump, D., Kiang, C.-L. and Berk, P. D. (1998) Hepatology **27**, 1064–1074
8 Berk, P. D., Wada, H., Horio, Y., Potter, B. J., Sorrentino, D., Zhou, S.-L., Isola, L. M., Stump, D., Kiang, C.-L. and Thung, S. (1990) Proc. Natl. Acad. Sci. U.S.A. **87**, 3484–3488
9 Stump, D. D., Zhou, S.-L. and Berk, P. D. (1993) Am. J. Physiol. **265**, G894–G902
10 Zhou, S.-L., Stump, D., Sorrentino, D., Potter, B. J. and Berk, P. D. (1992) J. Biol. Chem. **267**, 14456–14461
11 Zhou, S.-L., Stump, D., Kiang, C.-L., Isola, L. M. and Berk, P. D. (1995) Proc. Soc. Exp. Biol. Med. **208**, 263–270
12 Isola, L. M., Zhou, S.-L., Kiang, C.-L., Stump, D. D., Bradbury, M. W. and Berk, P. D. (1995) Proc. Natl. Acad. Sci. U.S.A. **92**, 9866–9870
13 Berk, P. D. and Stump, D. D. (1999) Mol. Cell. Biochem. **192**, 17–31
14 Obaru, K., Nomiyama, H., Shimada, K, Nagashima, F. and Morino, Y. (1986) J. Biol. Chem. **261**, 16976–16983
15 Tsuzuki, T., Obaru, K., Setoyama, C. and Shimada, K. (1987) J. Mol. Biol. **198**, 21–31
16 Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990) J. Mol. Biol. **215**, 403–410
17 DeLorenzo, R. J. and Ruddle, F. H. (1970) Biochem. Genet. **4**, 259–273
18 Berk, P. D., Delavega, L. and Stockert, R. J. (1998) Hepatology **28**, 400A (Abstract)
19 Soltys, B. J. and Gupta, R. S. (1996) Exp. Cell Res. **222**, 16–27
20 Sohocki, M. M., Sullivan, L. S., Harrison, W. R., Sodergren, E. J., Elder, F. F. B., Weinstock, G., Tanase, S. and Daiger, S. P. (1997) Genomics **40**, 247–252
21 Wood, T. C., Her, C., Askoy, I., Otterness, D. M. and Weinsilboum, R. M. (1996) J. Steroid Biochem. Mol. Biol. **59**, 467–478
22 Karinch, A. M. and Floros, J. (1995) Am. J. Respir. Cell Mol. Biol. **12**, 77–88
23 Stubbins, M. J., Harries, L. W., Smith, G., Tarbit, M. H. and Wolf, C. R. (1996) Pharmacogenetics **6**, 429–436
24 Bentley, D. R. and Dunham, I. (1995) Curr. Opin. Genet. Devel. **5**, 328–334
25 Gerhold, D. and Caskey, C. T. (1996) BioEssays **12**, 973–981
26 Kirkness, E. F. (1996) in Essays in Biochemistry, vol. 31 (Apps, D. K., ed.), pp. 1–9, Portland Press, London
27 Wand, D. G., Fan, J. B., Siao, C. J., Berno, A., Young, P., Sapolsky, R., Ghandour, G., Perkins, N., Winchester, E., Spencer, J. et al. (1998) Science **280**, 1077–1082


Received 2 September 1999; accepted 10 November 1999