

## Identification of the $\alpha$ -Amino adipic Semialdehyde Synthase Gene, Which Is Defective in Familial Hyperlysinemia

Katherine A. Sacksteder,<sup>1</sup> Barbara J. Biery,<sup>2</sup> James C. Morrell,<sup>1</sup> Barbara K. Goodman,<sup>3</sup> Brian V. Geisbrecht,<sup>1</sup> Rody P. Cox,<sup>4</sup> Stephen J. Gould,<sup>1</sup> and Michael T. Geraghty<sup>2</sup>

Departments of <sup>1</sup>Biological Chemistry, <sup>2</sup>Pediatrics, and <sup>3</sup>Obstetrics and Gynecology, Johns Hopkins University School of Medicine, Baltimore; and <sup>4</sup>Division of General Internal Medicine, University of Texas Southwestern Medical Center, Dallas

The first two steps in the mammalian lysine-degradation pathway are catalyzed by lysine-ketoglutarate reductase and saccharopine dehydrogenase, respectively, resulting in the conversion of lysine to  $\alpha$ -amino adipic semialdehyde. Defects in one or both of these activities result in familial hyperlysinemia, an autosomal recessive condition characterized by hyperlysinemia, lysinuria, and variable saccharopinuria. In yeast, lysine-ketoglutarate reductase and saccharopine dehydrogenase are encoded by the *LYS1* and *LYS9* genes, respectively, and we searched the available sequence databases for their human homologues. We identified a single cDNA that encoded an apparently bifunctional protein, with the N-terminal half similar to that of yeast *LYS1* and with the C-terminal half similar to that of yeast *LYS9*. This bifunctional protein has previously been referred to as “ $\alpha$ -amino adipic semialdehyde synthase,” and we have tentatively designated this gene “AASS.” The AASS cDNA contains an open reading frame of 2,781 bp predicted to encode a 927-amino-acid-long protein. The gene has been sequenced and contains 24 exons scattered over 68 kb and maps to chromosome 7q31.3. Northern blot analysis revealed the presence of several transcripts in all tissues examined, with the highest expression occurring in the liver. We sequenced the genomic DNA from a single patient with hyperlysinemia (JJa). The patient is the product of a consanguineous mating and is homozygous for an out-of-frame 9-bp deletion in exon 15, which results in a premature stop codon at position 534 of the protein. On the basis of these and other results, we propose that AASS catalyzes the first two steps of the major lysine-degradation pathway in human cells and that inactivating mutations in the AASS gene are a cause of hyperlysinemia.

### Introduction

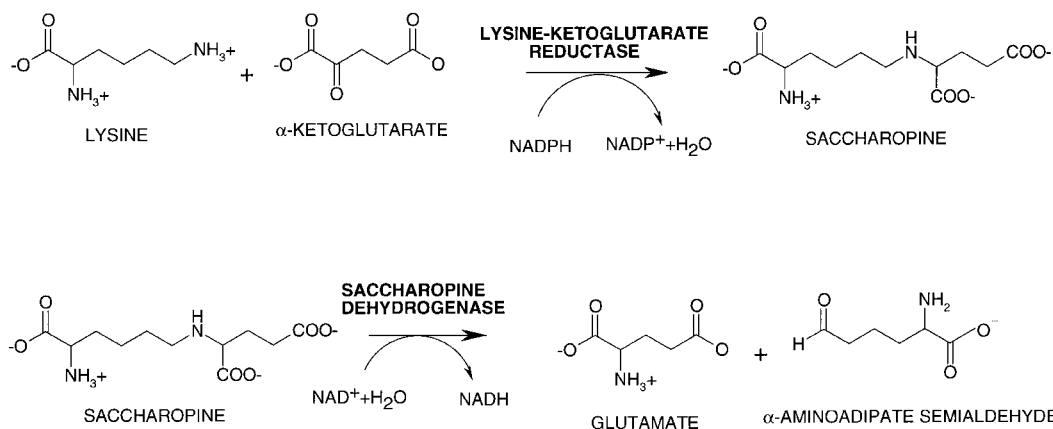
In humans, lysine is an essential amino acid, and there is no lysine biosynthetic machinery. However, humans do degrade lysine and saccharopine, and this catabolic pathway mirrors the terminal steps of lysine synthesis in lower eukaryotes. Lysine is first converted to saccharopine by the action of lysine-ketoglutarate reductase (LKR [E.C.1.5.1.7]). Saccharopine dehydrogenase (SDH [E.C.1.5.1.9]) then converts saccharopine to  $\alpha$ -amino adipate semialdehyde (fig. 1). In the lysine biosynthetic pathway of *Saccharomyces cerevisiae*, these two activities are encoded by two separate genes—*LYS1* and *LYS9*, respectively (Bhattacharjee 1985). However, in mammals, biochemical and genetic studies suggest the existence of a bifunctional protein that possesses both activities. Much of the data supporting the existence of

a bifunctional enzyme come from studies of familial hyperlysinemia (MIM 238700), a disorder that was first reported in individuals with physical and mental retardation (Woody 1964). However, subsequent studies identified pronounced hyperlysinemia in otherwise normal individuals, suggesting that hyperlysinemia alone may not be associated with a clinical phenotype (Ozalp et al. 1981). A biochemical analysis of these patients indicated that defects in LKR were the cause of the hyperlysinemia (Dancis et al. 1969). It was later noted that patients with these disorders often possessed a defect not only in LKR activity but also in SDH activity, suggesting the existence of a single locus encoding for both of these activities (Dancis et al. 1976). The related disorder saccharopinuria is also thought to be caused by defects in AASS (Cox and Dancis 1995). Additionally, biochemical studies of bovine liver revealed that LKR activity and SDH activity could not be purified away from each other, and that the protein migrated as a single band of 116 kD on SDS-PAGE analysis (Markovitz et al. 1984; Markovitz and Chuang 1987). We report here the identification and characterization of the human gene encoding  $\alpha$ -amino adipate semialdehyde synthase (AASS), the bifunctional protein that contains both LKR activity and

Received January 31, 2000; accepted for publication March 27, 2000; electronically published April 20, 2000.

Address for correspondence and reprints: Dr. Michael T. Geraghty, The McKusick-Nathans Institute of Genetic Medicine, Blalock 1008, The Johns Hopkins University School of Medicine, 600 North Wolfe Street, Baltimore, MD 21287-4922. E-mail: geraghty@jhmi.edu

© 2000 by The American Society of Human Genetics. All rights reserved. 0002-9297/2000/6606-0003\$02.00



**Figure 1** Human AASS, which catalyzes conversion of lysine to  $\alpha$ -amino adipic semialdehyde in the major lysine-degradation pathway

SDH activity. We demonstrate that mutations in AASS are responsible for the autosomal recessive disorder familial hyperlysinemia.

## Material and Methods

### Cloning of Human AASS

To identify the gene defective in hyperlysinemia, we used the BLAST algorithm to scan the human expressed-sequence-tag database (dbEST) for any sequences with the potential to encode proteins similar to either the LYS1 protein or the LYS9 protein. Three overlapping ESTs were identified as having similarity to LYS9, and the longest was obtained and sequenced. This partial cDNA clone contained, at its 5' end, a short open reading frame (ORF) that was homologous to a portion of LYS9. However, this ORF was too short to encode an enzyme with all the characteristics of LYS9, and we therefore concluded that this cDNA was lacking sequences at its 5' end.

Because there were no other clones for this cDNA in dbEST, we also scanned the nonredundant database in GenBank, for LYS1 and LYS9 homologues in other organisms. We identified in *Caenorhabditis elegans* a predicted bifunctional protein with an N-terminal domain similar to that of LYS1 and a C-terminal domain similar to that of LYS9. In addition, fragments of a *Mus musculus* homologue of LYS1 and LYS9 were also identified in this search. These protein sequences allowed us to identify conserved peptides between *C. elegans* and *M. musculus*, near the N-terminus of the LYS1 domain from these species. Degenerate oligonucleotides (5'-GA(A/G)GA(C/T)GT(A/C/G/T)AA(A/C/G)GC(A/C/T/G)TGGGA-3' and 5'-AA(C/T)GC(A/C/T/G)TGGGA(A/G)(A/C)G(A/C/T/G)CC(A/C/T/G)(C/T)T(A/C/T/G)GC(A/C/T/G)CC-3') encoding these peptides were then used in conjunction with a gene-specific oligonucleotide (5'-CC-

AGTCGACGTTGGTCATGAAGCTTTCCTATGG-3') from the 3' UTR of the human cDNA, to amplify a fragment of the human LYS1/LYS9 cDNA, in a pair of nested PCR reactions. The resulting PCR product was sequenced in its entirety. Further sequences at the N-terminus of LYS1 enzymes from other species led us to conclude that this fragment was likely to lack ~60 bp of coding region, plus whatever 5' UTR was present on the corresponding transcript(s).

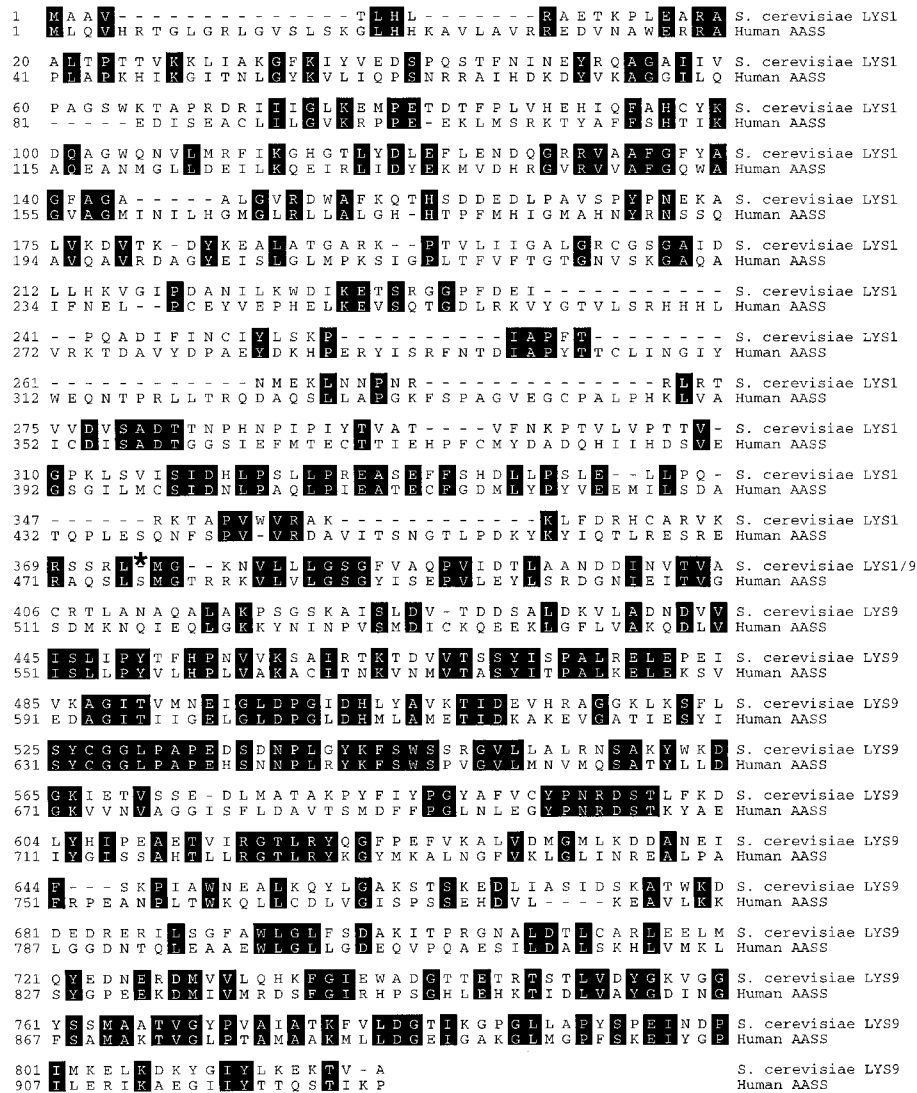
We employed a 5' RACE (rapid amplification of cDNA ends) procedure to identify additional sequences at the 5' end of this human cDNA. Marathon RACE human liver cDNA (Clontech) was used as template with the gene-specific antisense primer (5'-CAGGTAATGTACCGTTGGATG-3') and the adapter primer AP-1, in a PCR reaction. The longest PCR product was sequenced, and a full-length cDNA was assembled.

### Genomic Structure and Chromosomal Localization

The genomic structure was obtained by use of the full-length AASS cDNA to search the nonredundant nucleotide database, by means of the BLAST algorithm. A single human genomic DNA PAC clone, DJ1049N15, was identified (accession number AC006020). This 152,812-bp sequence contained segments that matched all parts of the AASS cDNA sequence and previously had been localized to chromosome 7q31.2-7q32. To confirm the chromosomal localization of AASS, we performed FISH, as described elsewhere (Sigurðardóttir et al. 1999). For this experiment, we used a mixture of amplified genomic DNA covering exons 7-11 (~3.2 kb) and exons 21-24 (~3 kb) as probes.

### Transcript Analysis and Mutation Analysis

Tissue-expression analysis was performed by standard protocols using multitissue northern blots from Clon-



\* indicates the end of the *S. cerevisiae* LYS1 protein and the beginning of *S. cerevisiae* LYS9

**Figure 2** Alignment of *S. cerevisiae* LYS1 and LYS9 proteins with the bifunctional human protein AASS. Yeast LYS1 protein is aligned with the N terminus of AASS, whereas yeast LYS9 protein is aligned with the C terminus of AASS. The junction between the yeast LYS1 and LYS9 proteins is indicated the asterisk (\*). Amino acid identities are shaded, and the amino acid positions are listed on the left. The N-terminus of AASS is 20% identical to that of *S. cerevisiae* LYS1, whereas the C-terminus of AASS is 38% identical to that of *S. cerevisiae* LYS9.

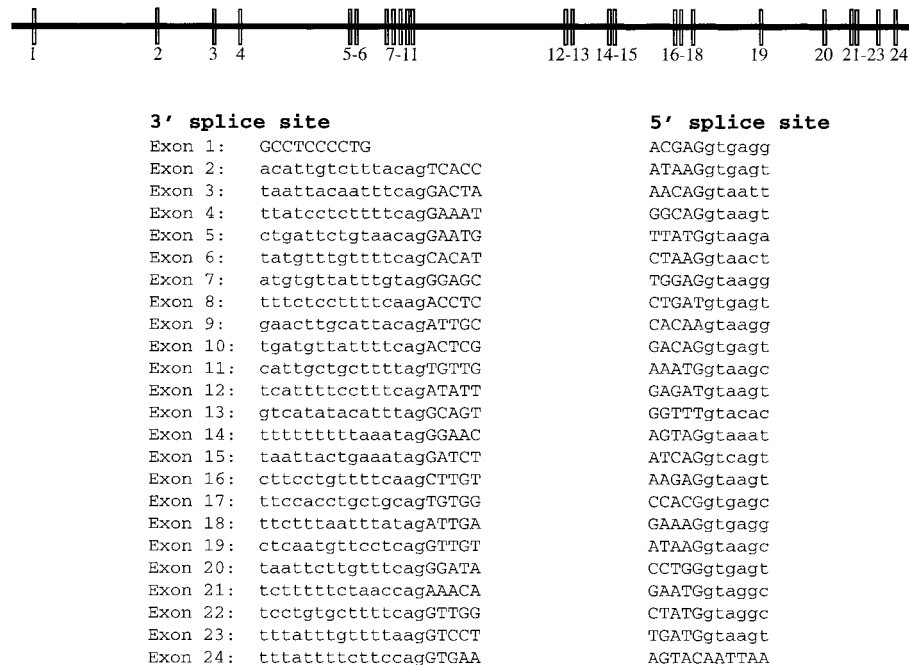
tech, with the entire AASS cDNA being used as a probe. For mutation analysis, genomic DNA was extracted from the fibroblasts of the patient, JJa (Dancis et al. 1976), and from wild-type fibroblasts (Coriell Cell Repositories, GM05756). We then amplified each exon of AASS from the two genomic DNA samples, using primers from the flanking intronic regions. Each PCR product was then gel-isolated and directly sequenced. For analysis of the patient's mRNA levels, we extracted total RNA from JJa and wild-type fibroblasts (Coriell Cell Repositories, GM05756 and GM08333). Northern blot analysis was then performed by standard protocols,

again with the full-length AASS cDNA being used as a probe. The same blots were probed with actin cDNA as a control.

**Results**

*Identification of AASS*

LKR and SDH are the first two steps in the lysine-degradation pathway in humans. Previous studies have suggested that hyperlysinemia and saccharopinuria are caused by a defect in a bifunctional protein with both



**Figure 3** Genomic structure of human AASS. At the top is a schematic representation of the genomic structure, showing the 24 exons of AASS, which span ~68 kb. Below this representation are the splice site junctions. All splice sites in AASS conform to the 5'...ag3' rule.

LKR activity and SDH activity (Dancis et al. 1976). In the present study, this bifunctional enzyme is referred to as "AASS." In the organism *S. cerevisiae*, which can synthesize as well as degrade lysine, the final two steps in the biosynthetic pathway are the reverse reactions of these two activities. Therefore, we used the two *S. cerevisiae* proteins, LYS1 and LYS9, to search the databases for similar human sequences. Using a variety of computer-based homology-probing approaches combined with molecular-biology techniques (reverse-transcription PCR, 5'-RACE, and DNA sequencing), we were able to derive an apparently full-length human AASS cDNA sequence. The cDNA sequence can be found in GenBank (accession number AF229180). This 2,987-bp cDNA contains a 2,781-bp ORF and is predicted to encode a 926-amino-acid-long protein with a molecular mass of 102 kD and an isoelectric point (pI) of 6.2. The N-terminal 374 amino acids display 20% amino acid identity to *S. cerevisiae* LYS1, whereas the C-terminal 552 amino acids display 38% sequence identity to *S. cerevisiae* LYS9 (fig. 2).

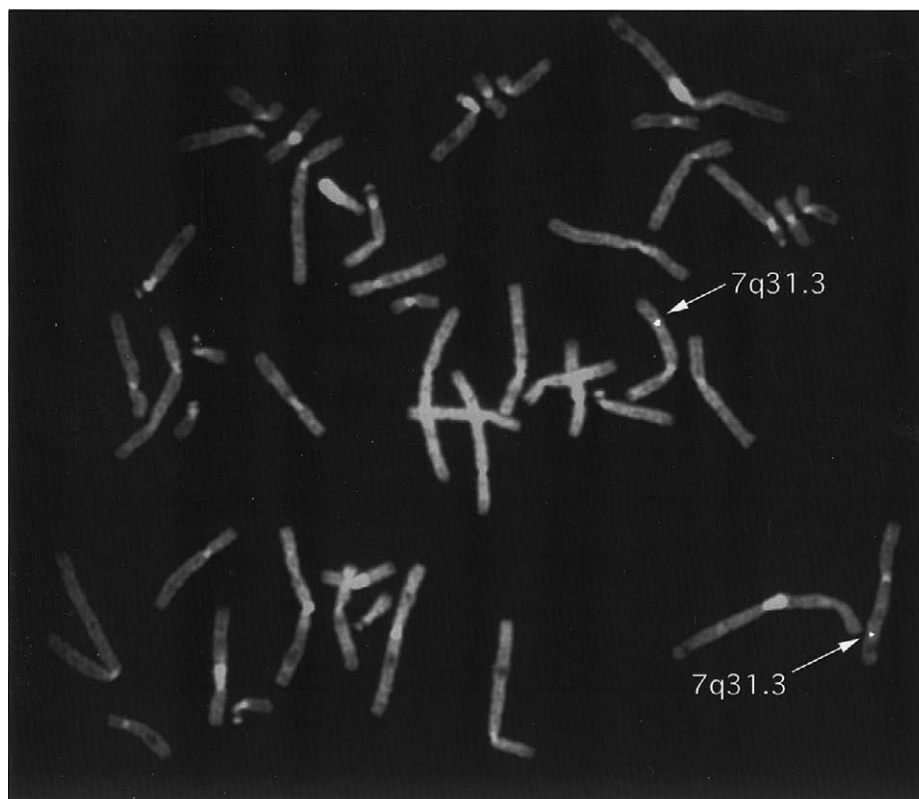
#### Genomic Structure and Chromosomal Localization of AASS

We used the BLAST algorithm to search the nonredundant database for any genomic sequence corresponding to that of the human AASS. The genomic clone DJ1049N15 contained the entire AASS gene, and, using

the BLAST two-sequences algorithm, we were able to deduce from this sequence the genomic structure of AASS. The AASS gene covers a region of ~68 kb and contains 24 exons (fig. 3). Exon 2 contains the first ATG of the ORF. All intron junctions conform to the 5'gt...ag3' consensus. The sizes of the exons range from 60 bp (exon 14) to 225 bp (exon 4). Introns varied in size from 96 bp to 11,400 bp, with an average length of 2,860 bp. Several sequence-tagged sites within the PAC containing the AASS gene had been mapped to the long arm of chromosome 7, between the markers D7S655 and D7S686, by radiation-hybrid mapping. To confirm this chromosomal localization by a different technique, we performed FISH. Human metaphase chromosomes were probed with the amplified genomic DNA, and a specific signal for the gene was observed at 7q31.3, a locus that corresponds well with the position determined by radiation-hybrid mapping (fig. 4).

#### Tissue-Specific Expression of AASS

The expression of the AASS mRNA in various tissues was analyzed by RNA hybridization. Northern blots containing poly(A)<sup>+</sup> RNA from 16 different tissue types were hybridized with a radiolabeled AASS-specific probe (fig. 5). Three transcripts were identified in all tissues examined. The major transcript was ~7 kb, with lesser bands at 9 kb and 3.4 kb. The 3.4-kb transcript is ap-



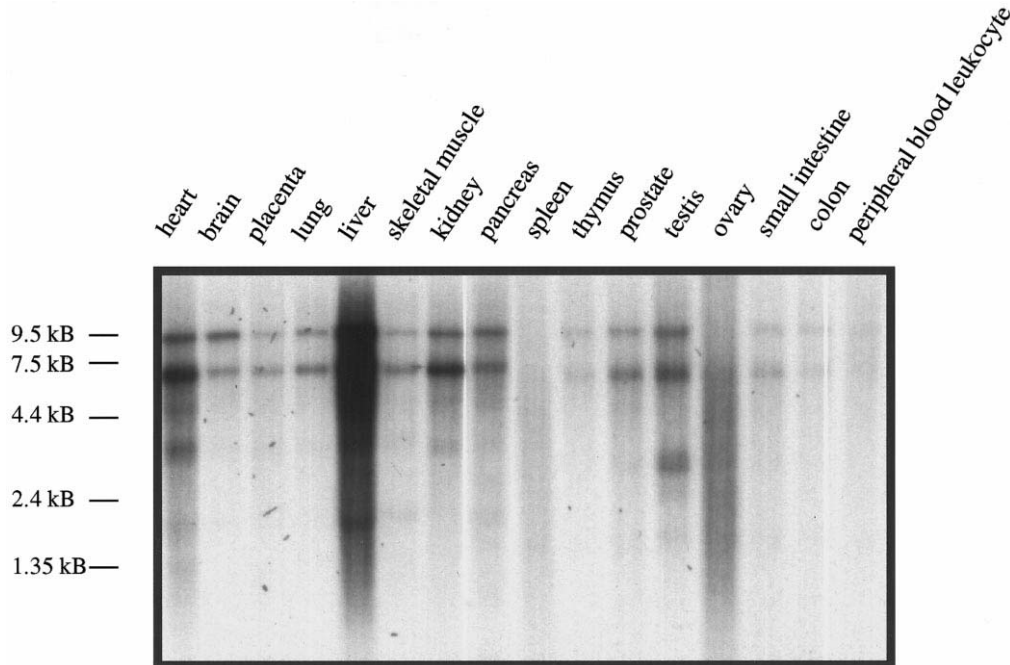
**Figure 4** Chromosomal localization of human AASS, by FISH. The arrow indicates the location of AASS on chromosome 7q31.3.

proximately the length predicted from our cDNA sequence after polyadenylation is taken into account. We currently do not know the molecular structure of the larger RNAs. However, the lack of any AASS homologues in the sequence databases suggests that they may be generated from the AASS gene reported here. In addition, we probed these same blots with the individual halves of the gene that correspond to the two activities. We found no change in the banding pattern, suggesting that these RNAs are all derived from the AASS gene. One possibility is that these transcripts correspond to incompletely spliced AASS heterogeneous nuclear RNAs. In support of this hypothesis, we did identify AASS ESTs that contained unspliced introns (GenBank accession numbers AA928584 and AA721427). The abundance of the AASS mRNA transcripts was varied, with the highest expression in the liver and with high expression in the heart and the kidney. The mRNA was at low but detectable levels in all other tissues.

#### *AASS Mutation in Hyperlysinemia*

A number of clinical studies have demonstrated that patients with hyperlysinemia are lacking in one or both of the enzymatic activities LKR and SDH. These patients have high levels of lysine in their plasma and urine, often

in conjunction with high levels of saccharopine. One such patient, JJa, has been described (Dancis et al. 1976). The product of a consanguineous marriage, JJa had severely reduced activities of both LKR and SDH, ~10% and ~3.5% of normal, respectively. We obtained a fibroblast cell line from JJa, prepared genomic DNA, and sequenced the exons of AASS. We identified a homozygous 9-bp deletion in exon 15 from this patient, but not from the control (fig. 6A), and we confirmed, by PCR analysis, that this deletion was not present in 100 control alleles (data not shown). This deletion is out of frame and results in a premature stop codon at the deletion junction at codon 534. This is predicted to reduce mRNA levels, resulting in reduced levels of both activities, a possibility that is supported by the finding of low activities of both LKR and SDH in this patient. To test this hypothesis, we isolated total RNA from the fibroblasts of JJa and from two unaffected fibroblast cell lines and performed a northern blot analysis, using the full-length AASS cDNA as a probe. We found that, although strong signals were detected for the AASS RNA from the unaffected cell lines, there was no detectable signal for the major 7-kb band in the RNA of JJa, and we found that the other signals were also severely reduced (fig. 6B).



**Figure 5** Tissue-specific expression of human AASS. A multitissue Northern blot (Clontech) was probed with the full-length human AASS.

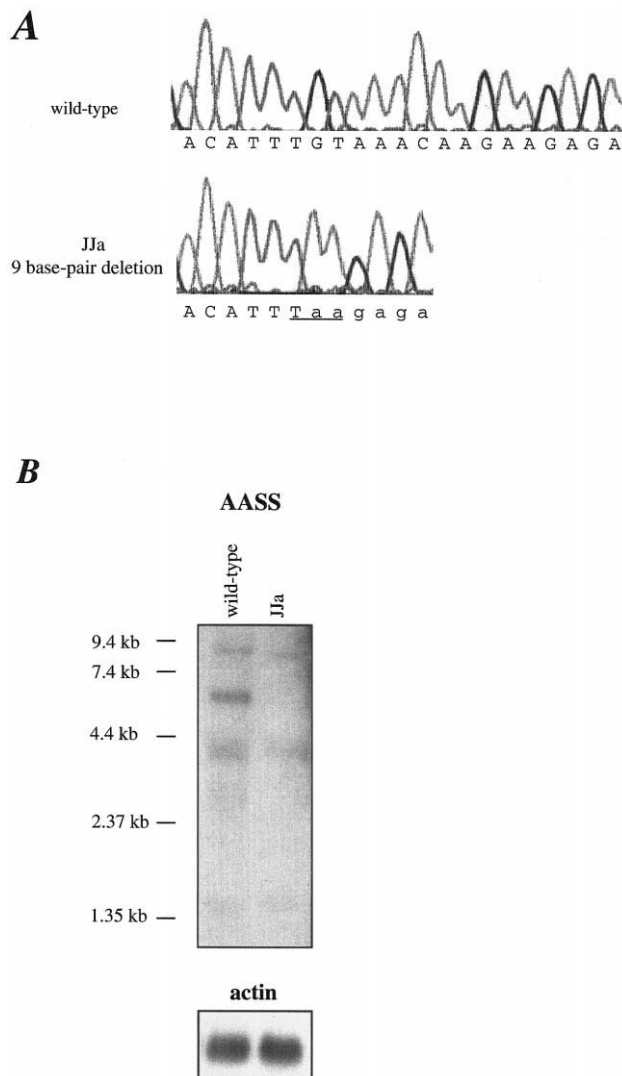
## Discussion

We report here the identification and characterization of a gene that encodes the human AASS protein. The N terminus of the predicted protein is 20% identical to that of the yeast LYS1, whereas the C terminus of the predicted protein is 38% identical to the yeast LYS9. In yeast, at least part of the lysine biosynthetic pathway is known to be peroxisomal, since two of the lysine biosynthetic enzymes, LYS1 and LYS4, terminate in peroxisomal targeting signals and have been demonstrated to be peroxisomal (Geraghty et al. 1999). The peroxisome also plays a known role in mammalian lysine metabolism. In addition to the major route of lysine metabolism, which degrades lysine via saccharopine, a minor route exists, which degrades lysine via pipecolate. This minor pathway is peroxisomal, and, although it is primarily used for D-lysine degradation, it is also the major pathway for L-lysine degradation in the brain (Chang 1978; Giacobini et al. 1980). These two pathways converge at amino adipic semialdehyde. The peroxisomal localization of both the yeast lysine biosynthetic pathway and the minor catabolic pathway in humans suggests the possibility that the major catabolic pathway in humans may also be peroxisomal. The peroxisomal signal type 1 (PTS1) is an obligately C-terminal tripeptide of sequence Ser-Lys-Leu (SKL<sup>\*</sup>) and variants (Gould et al. 1988). Both the mouse and human AASS enzymes terminate in sequences that may represent yet

another set of PTS1 variants, with the mouse AASS terminating in Ile-Lys-Leu (IKL<sup>\*</sup>) and with the human AASS terminating in the sequence Ile-Lys-Pro (IKP<sup>\*</sup>).

We found that the AASS gene includes 24 exons that span ~68 kb and is localized to chromosome 7q31.3. Analysis of the AASS transcripts revealed expression in all tissues examined, with highest expression in the liver. The liver is the site of the majority of amino acid metabolism, and, therefore, this finding is in agreement with the expected pattern for this enzyme.

In addition, we have identified a homozygous deletion in this gene in a patient with hyperlysinemia, a disorder that results from reduced activity of LKR and/or SDH. The 9-bp deletion is out of frame and results in the formation of a stop codon across the deletion, resulting in a predicted protein that lacks the 393 C-terminal amino acids and most of the putative SDH domain. Previous studies had demonstrated that this patient had only 3.5% of wild-type SDH activity. In addition, the LKR activity in this patient also was severely reduced, to only 10% of normal (Dancis et al. 1976). The reduction in the activities of both LKR and SDH in this patient suggests that this premature stop codon also affects AASS mRNA processing or stability. This theory is supported by our finding that fibroblasts from JJa had undetectable levels of the major AASS mRNA band and had lowered levels of the other bands. The related disorder saccharopinuria also has been suggested to result from mutations in the AASS gene. Pa-



**Figure 6** Mutation and expression analysis of AASS in hyperlysinemia. (A) Direct sequence analysis of an unaffected individual and JJa, a hyperlysinemic individual, demonstrating the 9-bp deletion in the AASS gene of JJa. The stop codon created by the out-of-frame deletion is underlined. The homozygous nature of the JJa sequence as well as the consanguineous nature of the parents suggests that the mutation is homozygous. (B) Northern blot of patient RNA and normal RNA. The top panel shows the blot probed with the AASS cDNA, while the lower panel shows the same blot probed with actin cDNA as a control.

tients with this disorder have both hyperlysinemia and saccharopinuria, although the saccharopinuria is much more severe. The few patients with saccharopinuria who have been studied suffer from neurological deficits, although the relation of the saccharopinuria to these symptoms is unclear (Cox and Dancis 1995). Although we were unable to obtain cell lines from patients with saccharopinuria, it seems likely that this less common

disorder results from specific mutations in the portion of AASS encoding SDH, which could explain the high levels of saccharopine, compared with that in patients with hyperlysinemia.

One of the interesting aspects of this metabolic pathway is the union of two single-function enzymes into one multifunctional enzyme. Lower eukaryotes such as *S. cerevisiae* have two distinct genes, one encoding for LKR activity and another encoding for SDH activity. Biochemical evidence previously presented had suggested the existence of a single gene encoding for the activities of both LKR and SDH in humans. We have confirmed this hypothesis, by the cloning of human AASS. Other higher eukaryotes, including *C. elegans* and plants (Epelbaum et al. 1997), also seem to have evolved a single-gene locus for these two activities. There could be a number of advantages to the joining of these two enzymatic activities into a single-gene locus. One possibility is the kinetic advantage of substrate channeling. By fusing the two activities together, the apparent  $K_M$  (Michaelis constant) for the substrates could decrease because the obstacle to diffusion has been overcome. A good example of this phenomenon is the bifunctional enzyme UMP synthase. This protein contains both orotate phosphoribosyltransferase activity and orotidine-5'-phosphate decarboxylase activity, and the substrate channeling of this enzyme is much higher than when the enzyme activities are in individual proteins (Pragobpol et al. 1984). Another possibility is that the fusion of the two activities into a single protein causes the protein to favor the degradation of lysine. This could explain the ability of *S. cerevisiae* to synthesize lysine with its two separate proteins, LYS1 and LYS9, whereas the bifunctional AASS serves to catalyze lysine degradation only. Future studies will allow us to elucidate the localization of lysine metabolism within the cell, as well as to assess the effects of bifunctionality versus monofunctionality for these two enzyme activities, which might suggest an evolutionary reason for this change.

## Acknowledgments

This work was supported in part by the National Institutes of Health grants DK45787 and HD10981 to S.J.G.

## Electronic-Database Information

Accession numbers and URLs for data in this article are as follows:

BLAST, <http://www.ncbi.nlm.nih.gov/BLAST>  
 Coriell Cell Repositories, <http://locus.umdj.edu/ccr/>  
 dbEST, <http://www.ncbi.nlm.nih.gov/dbEST>  
 GenBank, <http://www.ncbi.nlm.nih.gov/Genbank/Genbank>

Overview.html (for AASS cDNA [accession number AF229180] and AASS ESTs [accession numbers AA928584 and AA721427])

Online Mendelian Inheritance in Man (OMIM), <http://www.ncbi.nlm.nih.gov/omim> (familial hyperlysinemia [MIM 238700])

## References

- Bhattacharjee JK (1985) Alpha-aminoacidopate pathway for the biosynthesis of lysine in lower eukaryotes. *Crit Rev Microbiol* 12:131–151
- Chang YF (1978) Lysine metabolism in rat brain: the pipecolic acid-forming pathway. *J Neurochem* 30:347–354
- Cox RP, Dancis J (1995) Errors of lysine metabolism. In: Scriver CR, Beaudet AL, Sly WS, Valle D (eds) *The metabolic and molecular bases of inherited disease*. McGraw-Hill, New York, pp 1233–1238
- Dancis J, Hutzler J, Cox RP, Woody NC (1969) Familial hyperlysinemia with lysine-ketoglutarate reductase insufficiency. *J Clin Invest* 48:1447–1452
- Dancis J, Hutzler J, Woody NC, Cox RP (1976) Multiple enzyme defects in familial hyperlysinemia. *Pediatr Res* 10: 686–691
- Epelbaum S, McDevitt R, Falco SC (1997) Lysine-ketoglutarate reductase and saccharopine dehydrogenase from *Arabidopsis thaliana*: nucleotide sequence and characterization. *Plant Mol Biol* 35:735–748
- Geraghty MT, Bassett D, Morrell JC, Gatto GJ, Bai J, Geisbrecht BV, Hieter P, et al (1999) Detecting patterns of protein distribution and gene expression *in silico*. *Proc Natl Acad Sci USA* 96:2937–2942
- Giacobini E, Nomura Y, Schmidt-Glenewinkel T (1980) Pipecolic acid: origin, biosynthesis and metabolism in brain. *Cell Mol Biol* 26:135–146
- Gould SJ, Keller G-A, Subramani S (1988) Identification of peroxisomal targeting signals at the carboxy-terminus of four peroxisomal proteins. *J Cell Biol* 107:897–905
- Markovitz PJ, Chuang DT (1987) The bifunctional aminoacidic semialdehyde synthase in lysine degradation: separation of reductase and dehydrogenase domains by limited proteolysis and column chromatography. *J Biol Chem* 262: 9353–9358
- Markovitz PJ, Chuang DT, Cox RP (1984) Familial hyperlysinemias: purification and characterization of the bifunctional aminoacidic semialdehyde synthase with lysine-ketoglutarate reductase and saccharopine dehydrogenase activities. *J Biol Chem* 259:11643–11646
- Ozalp I, Hasanoglu A, Tuncbilek E, Yalaz K (1981) Hyperlysinemia without clinical findings. *Acta Paediatr Scand* 70: 951–953
- Pragobpol S, Gero AM, Lee CS, O'Sullivan WJ (1984) Orotate phosphoribosyltransferase and orotidylate decarboxylase from *Crithidia luciliae*: subcellular location of the enzymes and a study of substrate channeling. *Arch Biochem Biophys* 230:285–293
- Sigurðardóttir S, Goodman BK, Rutberg J, Thomas GH, Jabs EW, Geraghty MT (1999) Clinical, cytogenetic, and fluorescence in situ hybridization findings in two cases of “complete ring” syndrome. *Am J Med Genet* 87:384–390
- Woody NC (1964) Hyperlysinemia. *Am J Dis Child* 108:543