

L-Pipecolic acid oxidase, a human enzyme essential for the degradation of L-pipecolic acid, is most similar to the monomeric sarcosine oxidases

Gabriele DODT*, Do G. KIM†, Sylvia A. REIMANN*, Bernadette E. REUBER*, Katherine MCCABE†, Stephen J. GOULD‡ and Stephanie J. MIHALIK†‡¹

*Institut für Physiologische Chemie, Ruhr-Universität Bochum, 44780 Bochum, Federal Republic of Germany, †The Kennedy Krieger Institute, 707 N. Broadway, Baltimore, MD 21205, U.S.A., and ‡Departments of Pediatrics and Biological Chemistry, Johns Hopkins University School of Medicine, Baltimore, MD 21205, U.S.A.

L-Pipecolic acid oxidase activity is deficient in patients with peroxisome biogenesis disorders (PBDs). Because its role, if any, in these disorders is unknown, we cloned the associated human gene and expressed its protein product. The cDNA was cloned with the use of a reverse genetics approach based on the amino acid sequence obtained from purified L-pipecolic acid oxidase from monkey. The complete cDNA, obtained by conventional library screening and 5' rapid amplification of cDNA ends, encompassed an open reading frame of 1170 bases, translating to a 390-residue protein. The translated protein terminated with the sequence AHL, a peroxisomal targeting signal 1. Indirect immunofluorescence studies showed that the protein product was expressed in human fibroblasts in a punctate pattern that co-

localized with the peroxisomal enzyme catalase. A BLAST search with the amino acid sequence showed 31% identity and 53% similarity with *Bacillus* sp. NS-129 monomeric sarcosine oxidase, as well as similarity to all sarcosine oxidases and dehydrogenases. No similarity was found to the peroxisomal D-amino acid oxidases. The recombinant enzyme oxidized both L-pipecolic acid and sarcosine. However, PBD patients who lack the enzyme activity accumulate only L-pipecolic acid, suggesting that in humans *in vivo*, this enzyme is involved mainly in the degradation of L-pipecolic acid.

Key words: L-amino acid oxidase, fructosyl amino acid oxidase, organelles, peroxisome targeting signal 1, peroxisomes.

INTRODUCTION

The activity of L-pipecolic acid oxidase, an enzyme found in monkeys, rabbits and humans, is nearly absent from liver in patients with human peroxisome biogenesis disorders (PBDs) such as Zellweger syndrome [1,2]. In fact, the accumulation of its substrate L-pipecolic acid was one of the first biochemical abnormalities identified in these patients [3]. Functionally, the enzyme has been reported to be a component of the alternative lysine degradation pathway via L-pipecolic acid [4] that converts L-pipecolic acid to Δ^1 -piperidine-6-carboxylic acid. In monkeys and humans, the enzyme is an H₂O₂-forming peroxisomal oxidase [5] that contains covalently bound FAD [6]. Although the peroxisome contains at least two D-amino acid oxidases (D-amino acid oxidase and D-aspartate oxidase), no other L-amino or L-imino acid oxidase has been associated with the peroxisome [7].

Patients with PBDs have severe neurological dysfunction, are profoundly mentally retarded and have little peroxisomal enzyme function [7]. In addition, patients with another neurodegenerative disorder, glutaric aciduria type II, also accumulate L-pipecolic acid (R. I. Kelley, personal communication). The underlying biochemical aetiology of the neurological deficits is unknown. However, because both L-pipecolic acid and its metabolites are reported to have neural activity [8,9], these compounds might have a role in the pathophysiology of these disorders. It therefore

seems important to elucidate the enzymes associated with L-pipecolic acid degradation.

In the present study we took purified L-pipecolic acid oxidase from monkey liver and cut the protein into smaller peptide fragments; we then purified several internal peptides. These fragments were sequenced and the internal amino acid sequence was used in a reverse genetics approach to identify the human cDNA for this gene. Surprisingly, a comparison of the amino acid sequence of L-pipecolic acid oxidase with that of known proteins suggests that it is most closely related to the monomeric sarcosine oxidases rather than to the peroxisomal D-amino acid oxidases.

MATERIALS AND METHODS

Materials, cell lines, bacterial strains and libraries

Endoproteinase Glu-C was obtained from Boehringer Mannheim. Human liver RNA was a gift from Dr. Martina McGuinness. Anti-(L-pipecolic acid oxidase) was generated in rabbits by using purified protein, as described previously [10]. Sheep anti-(human catalase) was obtained from The Binding Site (San Diego, CA, U.S.A.). The human liver cDNA library was obtained from Dr. David Valle. Fibroblast cell lines were obtained from the Mental Retardation Research Center at Kennedy Krieger Institute and then transformed with simian virus 40.

Abbreviations used: PBD, peroxisome biogenesis disorder; PTS 1, peroxisome targeting signal 1; RACE, rapid amplification of cDNA ends; RT-PCR, reverse-transcriptase-mediated PCR.

¹ To whom correspondence should be addressed at the Kennedy Krieger Institute (e-mail mihalik@kennedykrieger.org).

The nucleotide sequence data reported will appear in DDBJ, EMBL and GenBank Nucleotide Sequence Databases under the accession number AF134593.

Table 1 Peptide sequences obtained from monkey liver L-pipecolic acid oxidase and associated oligonucleotides

Intact purified protein was cut with either CNBr or endoproteinase Glu-C and the resulting peptides were purified and sequenced as described in the Materials and methods section. The underlined region in peptide 1 represents the sequence used to construct the successful anti-sense degenerate oligonucleotide; the underlined region in peptide 3 was used to construct the sense degenerate oligonucleotide.

Peptide no.	Cutting method	Amino acid sequence	Sense/anti-sense	Degenerate oligonucleotide
1	CNBr	KENQELKTIQATLXRGRVEXQXL	A	GTNGCYTGDATNGTYTTNARYTCYTG
2	Endoproteinase Glu-C	LKTIQASLSRQRVE		
3	Endoproteinase Glu-C	<u>QFFLPHSRGSSHGQSRIIRKAYLE</u>	S	CARTTYTTYTNCNCAY

Determination of internal peptide sequences

L-Pipecolic acid oxidase was purified from monkey (*Macaca mulatta*) liver as described previously [10]. To obtain internal peptides, 0.5–1 mg of purified protein was cut directly with CNBr or with endoproteinase Glu-C after reaction with iodoacetamide as described by Matsudaira et al. [11], except that in this reaction 4 M guanidinium chloride in 0.1 M NH_4HCO_3 , pH 8.2, replaced urea and Tris buffer respectively. After incubation, the degraded protein was directly injected on a Vydac C_4 HPLC column equilibrated with 0.054% (v/v) trifluoroacetic acid in water (buffer A) and eluted with a gradient of 0–55% buffer B [50% (v/v) acetonitrile/0.06% (v/v) trifluoroacetic acid] over 60 min at a flow rate of 0.5 ml/min. Peptides were identified by their absorbance at 220 nm. Fractions of interest were freeze-dried, solubilized in 1% (v/v) trifluoroacetic acid and subjected to Edman degradation.

Generation of a DNA probe by reverse-transcriptase-mediated PCR (RT–PCR) with mixed oligonucleotides

Sequence from two of the peptides was used to construct four mixed oligonucleotides (Table 1) (Operon). The product of reverse-transcribed human liver RNA (GeneAmp RNA PCR Kit; Perkin-Elmer Cetus) was used as a template in a 100 μl PCR reaction. PCR was as described by Lee and Caskey [12] for degenerate oligonucleotides, with an annealing temperature of 37 °C for cycles 2–5 and 45 °C for the next 30 cycles. Reaction products were separated on a 3% (w/v) NuSieve (FMC, Rockland, ME, U.S.A.)/1% (w/v) agarose gel, subcloned into pUC19 to form pGD55, and sequenced. Control reactions were performed with each mixed oligonucleotide independently. They yielded several non-specific PCR products that could be eliminated from the results of dual oligonucleotide reactions.

Screening a human liver $\lambda\text{gt}11$ library

Bacteriophage were plated on *Escherichia coli* strain Y1090 at a concentration of 50000 phage per 150 mm plate on eight plates. Standard protocols were used to identify and isolate candidate clones [13]. The 180 bp RT–PCR product was cut from pGD55 and randomly primed with ^{32}P [13]. The filters were hybridized with the probe at 42 °C. After being washed, the filters were dried and subjected to autoradiography for 48 h. Positive plaques were eluted and rescreened. Plaques were judged as pure after three rounds of screening.

The phage DNA was amplified and purified as described previously [14]. The specific insert was cut with *EcoRI* and isolated from a 0.75% agarose gel. The isolated fragment was inserted into the phosphatase-treated [13] *EcoRI* site of pSKII to

make pGD60. The ligated DNA was transfected into competent cells (DH5 α) and the cDNA insert was sequenced.

5' rapid amplification of cDNA ends (5' RACE)

Clones containing the 5' end of the transcript were generated by using a 5' RACE Kit (Life Technologies). For reverse transcription, 1 μg of total RNA from human liver was mixed with 2.5 pmol of oligonucleotide Pip127 (5'-CCCGAATTCCAATTGGGTTCCAGCCTCGTG-3') derived from the sequence of the original clone. A nested internal oligonucleotide Pip86 (5'-GGCGAATTCGATAGCACTCATGCATCATC-3') and the anchor primer provided with the kit were used to prime the subsequent PCR reaction. PCR amplification was performed for 35 cycles for 45 s at 94 °C, followed by 25 s at 57 °C and 2 min at 72 °C. Potential products were separated on 3% (w/v) NuSieve/1% (w/v) agarose gels. Because no product was detected after this first round of PCR, 20 μl of this reaction was used as a template for a secondary PCR reaction (100 μl) under the same conditions. DNA obtained from PCR reactions was digested with *SalI* and *EcoRI* at the respective unique sites that had been incorporated into the primers. The fragments were cloned into pUC19 at the corresponding sites to form pGD63. PCR products were considered authentic if the sequences of regions overlapping the original clone were identical. To correct for *Taq* polymerase errors, clones from three different RT–PCRs were sequenced.

Preparation of the full-length cDNA clone

To insert the 5' end of the gene (in pGD63) into the original clone (pGD60), both plasmids were cut with *SalI* and *BbsI*. After gel purification, the PCR-derived 5' fragment from pGD63 was ligated into pGD60 to yield pGD65.

Expression of L-pipecolic acid oxidase clones in *E. coli*

To form a recombinant gene for bacterial expression studies, pGD60 was cut with *Pvu2* (at bp 29 of the insert) and *EcoRI*, whereas pMAL-c2 (New England Biolabs) was cut with *XmnI* and *EcoRI*. Appropriate fragments were purified and cloned as described previously [14], resulting in pMALc2PIPOX98-390 (in which PIPOX stands for pipecolic acid oxidase). This plasmid, and a pMALc2 plasmid lacking an insert, were transformed into DH10B and induced with 0.3 mM isopropyl β -D-thiogalactoside for 4 h at 37 °C. Bacterial pellets were separated on SDS/4–20% (w/v) polyacrylamide gradient gels (Bio-Rad) by the method of Laemmli [15]. When appropriate, proteins were transferred to nitrocellulose and blots were probed with crude anti-(L-pipecolic acid oxidase) [10] (1:500 dilution) as described by O'Farrell [16]. Goat anti-rabbit secondary antibodies with horseradish peroxidase were used with enhanced chemiluminescence (Amersham) to detect the immunoreactive materials.

For the large-scale production of enzymically active protein, a second pMAL-c2 clone (pMALc2-PIPOXt) containing the entire open reading frame was constructed by cutting pGD65 at *KpnI*, followed by blunting of the cut ends with Klenow enzyme (Boehringer Mannheim) and a subsequent cut with *EcoRI*. The fragment containing the open reading frame was then inserted into the *XmnI* and *EcoRI* sites of pMAL-c2. After electroporation into DH10B, the protein was induced in 0.3 mM isopropyl β -D-thiogalactoside at 30 °C for 7 h and lysed as described previously [14]. The recombinant protein was affinity-purified from the soluble bacterial fraction by using an amylose column in accordance with the manufacturer's directions (New England Biolabs).

Sequence analysis

Multiple alignments of amino acid sequences and calculation of percentage amino acid identity and similarity were performed with the Clustal W program [17]. The phylogenetic tree was generated with the Phylogenetic Analysis Using Parsimony Program (PAUP, version 4.0d64, kindly provided by Dr. David Swofford, Smithsonian Institute, Washington, DC, U.S.A.).

Transient transfection and immunofluorescence

A *Sall* and *SpeI* fragment from pGD65 was cloned into pcDNA3 (Invitrogen) at *XhoI* and *XbaI* respectively to form pSM3. After purification, the plasmid was transfected into transformed human fibroblasts by using Lipofectamine (Life Technologies) in accordance with the manufacturer's directions. After 44–48 h the transfected cells were fixed and subjected to indirect immunofluorescence as described previously [18], with affinity-purified rabbit anti-(human L-pipecolic acid oxidase) and sheep anti-(human catalase).

Northern blot

The full-length clone was cut from pSM3 at *BamHI* and *XbaI*, gel-purified and randomly primed with [³²P]CTP with the Rediprime kit (Amersham). The probe was purified on a Nuc Trap column (Stratagene) and was hybridized to a prepared Human Multiple Tissue Northern blot (Clontech) with ExpresHyb (Clontech) by following their protocol. The signal was detected with a Fuji PhosphorImager.

Determination of enzyme activity

Oxidase activity was determined by horseradish peroxidase-linked fluorometric assay exactly as described previously [14], with either sarcosine or L-pipecolic acid as substrates. Protein concentration was determined by the method of Bradford [19], with BSA as standard.

RESULTS

Resolution and amino acid sequencing of three peptides obtained from purified L-pipecolic acid oxidase

At first we found that the N-terminus of L-pipecolic acid oxidase purified from *Macaca mulatta* was blocked to amino acid sequencing. Subsequently the purified protein was cut with both CNBr and endoproteinase Glu-C and three relevant internal peptides were HPLC purified and sequenced (Table 1). Peptides 1 and 2, obtained by CNBr and Glu-C cutting respectively, overlapped. Portions of these sequences (underlined in Table 1) were used to design four degenerate oligonucleotide primers (Table 1). PCR was performed with human liver RNA using both relevant combinations of oligonucleotides. One 180 bp product

```

c att tgc att tga ggc tct grg got gtg ggg ctg agt ggc atc
ATG GGG GCT CAG AAA GAT CTC TGG CAC GCC ATT CTG ATT GGG GGG GGG ATC CAG GGC TGC 60
M A A Q K D L W D A I V I G A G I Q G C
TTC ACT GCA TAC CAC CTC GCC AAA CAC AGG AAG AGG ATT CTC CTC CTG GAG CAG TTC TTT 120
F I A Y H L A K H R K R I L L L E Q F F
CTA CCA CAC TCC CGA GGA AGC TCC CAT GGA CAA AGC CGG ATA ATC CGA AAG GCG TAC CTG 180
L P H S R R G S S H G Q S R I I R K A Y L
GAA GAC TTT CAC ACC CGG ATG ATG CAT GAG TGC TAT CAG ATA TGG GGC CAG CTG GAG CAC 240
E D F F Y C R M M H E C Y Q F W A Q L E H
GAG GCT GGA ACC CRA TGG CAC AGG CAG ACT GSA TTA CTG CTC CGG GAA RTG AAA GAG AAT 300
E A G T Q L H R Q T S L L L L L S H K R N
GAA GAA TTA AAG ACA ATC CAG GCC AAT CTG TCG AGC CAG ASG GTA GAA CAG CAG TGT CTT 360
C E L K T I Q A N L S R Q R V E H Q C L
TCA TCT GAG GAA CTG AAG CRA CGT TPC CCA AAC ATT CCG TTG CCC AGG GCA GAA CTG GGG 420
S S R E L K Q R P P N I R L P R G E V G
CCC TTG GAC AAT TCC GGA GGA GTT ATC TAT GCA TAC AAG GCC CTC AGA GCC CTG CAG GAT 480
L L D N S G G V I Y A Y K A L R A L Q D
GCA ATT CGA CAG CTA GGA GGC ATA GTG CBT GAC GGA GAG AAG GTG GTG GAG CTA AAC CCA 540
A I R Q L G G I V R D G E K V V E I N P
GGG CTA CTG GTC ACG GTG AAA ACC ACC TCC AGG AGC TAC CAA GCG AAG AGC TTG GTC ATC 600
G L I V T V K T T S S Y Q A K S I V I
ACA GCA GGT TCT TGG ACC AAC CAG CTC CTC CGT CCC CTG GGC ATT GAG ATG CTT CTC CAG 660
T A G P R T Y Q L L R P L G E M F L Q
ACC CTG CCG ATC AAC GTC TGT TAC TGG GGA GAG ATG GTT CCT CGG AGT TAT GAT GTG TCC 720
T L R I N V C Y W R E M V P G V N
CAG GGC TTT CCG TGC CTC TGG CTG GGC TTG TGT CCC CAC CAC ATC TAC GSA CTG CCC 780
Q A F P C F L W L G L C P H H I Y G L P
ACA GSA GAG TAC CCA GGG CTG ACG AAG GTC AGC TAT CAC CAC GGC AAC CAC GGA GAC CCT 840
T G E Y P G L M K V S Y H H G N H A D P
GAG GAG CCG GAC TGC CCC ACA GCA CGC ACA GAT CTA GCA GGC CAG CTG ATC AGC AGC 900
E E R D C P T A R T D I S D V Q I L S S
TCT STC AGA GAT CAC TTA CCT GAT CTG AAG CCC GAG CCT GCT GTC ATT GAG AGC TGC ATG 960
F V L H L F D L K P E P A V I E S C M
CAC ACG AAT ACC CCT GAT GAG CAG TTC ATT CTC GAT CCG CCA CCA AAG TAT GAC AAC ATT 1020
Y T N T P D E Q F I L D R H F K Y D N I
GTC ATY GGT GCT GGA TTC TCT GGG CAC GGG TCT AAG CTG GGC CCT GCT GTG GGG AAG ATC 1080
I G A C F S G H G F K L A F V V G K I
CTG TAT GAA CTA AGC ATG AAA TTA ACA CCA TCT TAT GAC TAT GCA CTT GTC ACA ATC GCA 1140
L Y E L S M K L T P S Y D L A F F R I S
CGT TTC CCA AGC CTG GCC AAA GGC CAC CTG TGA 1173
R F F S L A K A H L STOP
cctctggccagaagcctccctctctgtgcacgggagccagttccacagatgggagaagatg:ctcagatgaaggagtgctccctcg
agatacatcctttcttctgctgcctgaatccccataaacaccagctgagctacaccttcttctctggccctctcc
cttttctctgctcactggaatccccgtaaacaccagacgatggagctgaccttcttctctgctccactccccat:ca
aactacttctctgctccccgaagtgatcagatattctacaatcacagagtagaagaagcgtctgagatggg:atatacagtaa
gaagagggcacaagaactgctctggagacccaaagcagctgtataaat:tcctcttctgggggcccgtgtggggt:caac
ctgtaalctcgggaccttgggggctggagacagcagga:caactgagccaggaat:cctgagcccg:gaattcaagccacagc
ctggcccaactcgtgaaacccatctctccactaaaaatacaaaatagctgggtgtggtggagatgct:gtaatccccagcta
cttgggagactgggcaaggaactcaactgaaaccccggaggtggaggtggcagtgagcagtgagccagagatgagccattcgaactccag
ctctgggcaaaaggtggagcctctctcaaaaaaaacacaaacacccctcttggcggatlatc:gaagagctctgg
gatgatttgac:aaactccccggaggggtgagaaaggtgactcaagaat:caactatgccccttggaggt:cttctccct
aaagctaaagcagcctccatcaaaatgcccaataaaatctggctgctcttggagcaacataaatatttgggaacaaata
aataactgcctctgcaaaaaaasaaazcaaaaaaaaxxxxxxaaaaaaaggaattc

```

Figure 1 Nucleotide sequence of the L-pipecolic acid oxidase cDNA from human liver and the amino acid sequence of its corresponding gene product

The coding region is denoted by capital letters; untranslated regions are denoted by lower-case letters. The original gene probe generated from the degenerate oligonucleotides is underlined, as is the C-terminal PTS 1 targeting signal AHL. The original amino acids identified from the fragments of native protein are indicated in bold.

(Figure 1, underlined) was produced with the anti-sense degenerate oligonucleotide from peptide 1 and the sense oligonucleotide from peptide 3. Products were considered authentic when their sequence translation matched that of the regions of the original peptides that were not incorporated into the degenerate oligonucleotides (see Figure 1 and Table 1).

Cloning of cDNA for full-length L-pipecolic acid oxidase from human liver

A λ gt11 cDNA library from human liver was screened with the 180 bp probe and one clone of 1824 base pairs was obtained. When this phage fragment was cut with *EcoRI*, it was found to have an internal cutting site 468 bp from the terminus of the 3' untranslated region. After sequencing, the 5' fragment was found to encompass an incomplete open reading frame of 1014 bp that encoded residues 53–390 of the final protein. A partial protein product (residues 98–390 of the final protein) from this clone was induced in *E. coli* as a recombinant DNA product of the expression vector pMAL-c2. This protein product (69 kDa) interacted with affinity-purified anti-(L-pipecolic acid oxidase) (obtained from native protein) on a Western blot (Figure 2). The chimaeric product was slightly smaller than expected (40 kDa for the maltose-binding protein and 32 kDa for the gene-specific product). No band was identified in the control lane, in which the maltose-binding protein alone was loaded. Taken together, these

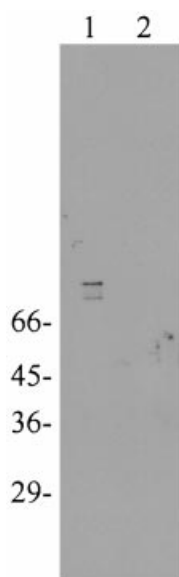


Figure 2 Western blot of the protein product of the original human L-pipecolic acid oxidase clone

The original cDNA encoding residues 98–390 of L-pipecolic acid oxidase was cloned into pMAL-c2 and transformed into bacteria; the protein product was expressed as described in the Materials and methods section. A protein preparation from a bacterial strain containing the pMAL-c2 construct alone was also prepared. Both protein products were separated on an SDS/10% (w/v) polyacrylamide gel and blotted as described with polyclonal anti-(L-pipecolic acid oxidase) generated from the purified native protein derived from monkey liver. Lane 1 contained the product of the recombinant pMAL-c2 clone; lane 2 contained the pMAL-c2 product alone.

results suggested that this single clone coded for much, but not all, of the expected gene product of 45 kDa [10].

To obtain the missing 5' portion of the gene, 5' RACE was performed with human liver RNA. The product obtained from this reaction was subcloned into pUC19 and three different clones were sequenced. All contained the same nucleotide sequence.

The complete cDNA for L-pipecolic acid oxidase (Figure 1) contained a 43 bp 5' untranslated region with an open reading frame of 1170 bp and a 3' untranslated region of 998 bp. The open reading frame encodes a protein, L-pipecolic acid oxidase, encompassing 390 residues with a molecular mass of 44 kDa. Structural analysis of the translation product showed that the N-terminus contains a sequence consistent with an ADP- $\beta\alpha\beta$ -binding fold that associates nucleotide cofactors such as FAD and NAD⁺ with proteins. Furthermore, the protein product terminates with the tripeptide AHL (Figure 1), an identified mammalian peroxisome targeting signal 1 (PTS 1) [20].

Human L-pipecolic acid oxidase is most similar to the monomeric sarcosine oxidases

Analysis of the complete amino acid sequence by using BLAST algorithms [21] showed that the amino acid sequence was most similar to rabbit sarcosine oxidase (accession no. OCU82267) [14] and to a mouse clone for sarcosine oxidase (accession no. U94700) [22], with the next greatest similarity to an unidentified *Caenorhabditis elegans* gene product (accession no. U23529). These results were not surprising because our original L-pipecolic acid oxidase clone was used as a probe to screen for the rabbit gene [14] and the *C. elegans* gene was first identified during a



Figure 3 Sequence alignment between human L-pipecolic acid oxidase, rabbit sarcosine oxidase (SOXp) and the monomeric sarcosine oxidase from *Bacillus* sp. NS-129

After the gene products with greatest similarity to L-pipecolic acid oxidase had been identified by BLAST searching, they were aligned with Clustal by using a PAM250 residue weight table. Identical amino acids are denoted by an asterisk; conserved amino acids are denoted by a colon. The sequence regions overlined and numbered 1–4 are regions common to all monomeric sarcosine oxidases. Region 1 contains the ADP- $\beta\alpha\beta$ -binding fold.

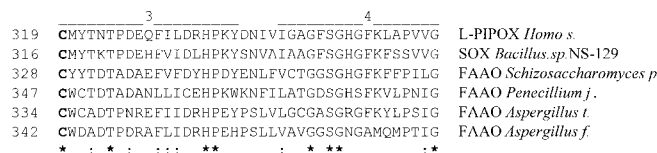


Figure 4 Sequence alignment between human L-pipecolic acid oxidase, the monomeric sarcosine oxidase from *Bacillus* sp. NS-129 and the fructosyl amino acid oxidases in the commonly conserved regions 3 and 4

Amino acid sequences in L-pipecolic acid oxidase and the fructosyl amino acid oxidases corresponding to regions 3 and 4 (overlined) of the monomeric sarcosine oxidases were aligned with Clustal by using a PAM 250 residue weight table. Identical amino acids are denoted by an asterisk; conserved amino acids are denoted by a colon. The conserved cysteine residue essential for monomeric sarcosine oxidase enzyme activity is shown in bold. The full names of the proteins with their sources are as follows: L-pipecolic acid oxidase (L-PIPOX), *Homo sapiens*; sarcosine oxidase (SOX), *Bacillus* sp. NS-129; fructosyl amino acid oxidase (FAAO), *Schizosaccharomyces pombe*; fructosyl amino acid oxidase, *Penicillium janthinellum*; fructosyl amino acid oxidase, *Aspergillus terreus*; fructosyl amino acid oxidase, *Aspergillus fumigatus*.

BLAST search with the rabbit sequence. As shown in Figure 3, the rabbit and human proteins differed by only three residues, suggesting that the protein is very conserved. However, L-pipecolic acid oxidase also showed a 31% amino acid identity and a 53% similarity to the *Bacillus* sp. NS-129 monomeric sarcosine oxidase. The areas of highest similarity among the proteins included an N-terminal ADP- $\beta\alpha\beta$ -binding fold (underlined region 1, Figure 3) [23], as well as three other regions that are conserved in all four of the bacterial monomeric sarcosine oxidases so far identified [14]. Region 1 is common to all sarcosine oxidases and dehydrogenases.

The BLAST search identified another protein family, the fungal fructosyl amino acid oxidases, as showing similarity to L-pipecolic acid oxidase, albeit to a lesser degree. When the family was further investigated by alignment with both L-pipecolic acid

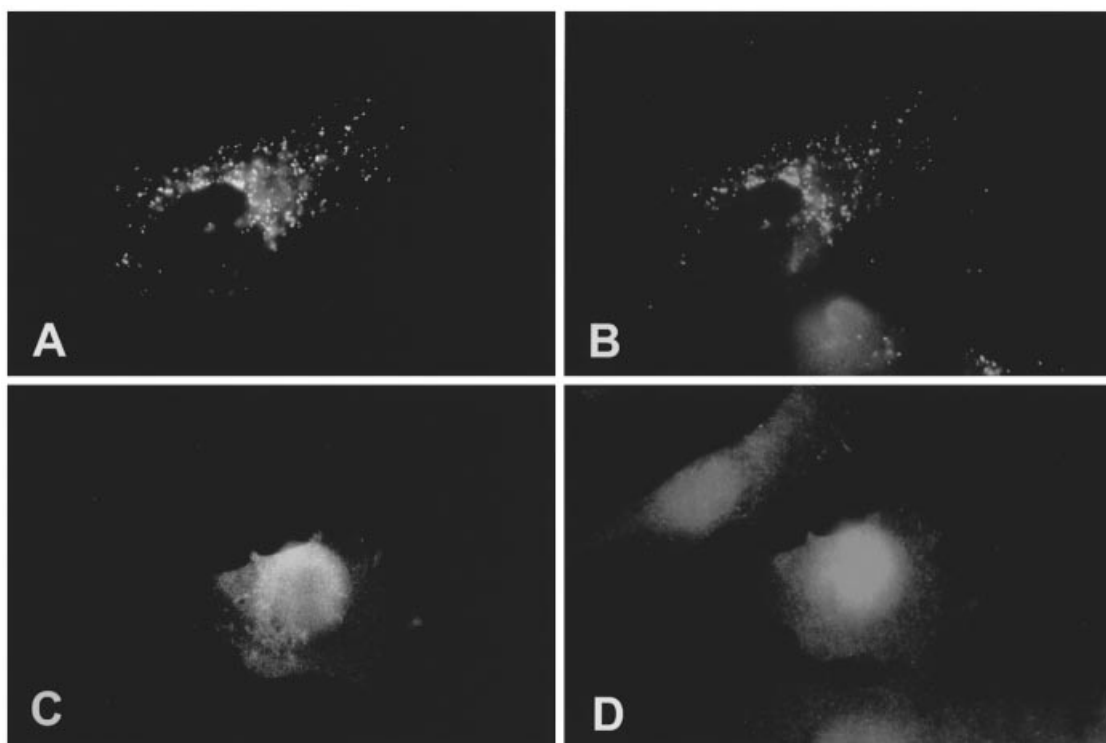


Figure 5 Subcellular localization of human L-pipecolic acid oxidase in normal transformed fibroblasts and in fibroblasts from patients with defective PTS 1 protein import

Normal fibroblasts (GM5756) (A, B) and 005-T (deficient import of PTS 1 proteins) (C, D) were transfected with pSM 3 in the presence of Lipofectamine. For indirect immunofluorescence, the cells were incubated with rabbit anti-(human L-pipecolic acid oxidase) and sheep anti-(human catalase), followed by incubation with secondary antibodies conjugated FITC and tetramethylrhodamine isothiocyanate. Both L-pipecolic acid oxidase (A) and catalase (B), the marker for PTS 1 proteins, were imported into peroxisomes, whereas neither was imported into peroxisomes of the PTS 1 receptor-defective patient (C, D). The originals were processed from colour negatives to monochrome prints with Adobe Photoshop 4.0.

oxidase and a representative monomeric sarcosine oxidase (*Bacillus* sp. NS-129), the members of this family were only 12–15% identical with these proteins. However, the homologous regions in the fructosyl amino acid oxidases were limited to homologous regions 3 and 4 (Figure 4), as well as to the region 1, an area common to all sarcosine-degrading enzymes.

Expressed chimaeric product of maltose-binding protein and L-pipecolic acid oxidase can utilize both L-pipecolic acid and sarcosine as substrates

To determine whether the cloned gene product functions as an L-pipecolic acid oxidase, the full open reading frame was cloned in-frame downstream of maltose-binding protein; its protein product was then expressed and affinity purified. Derived K_m values were 1.9 and 6.7 mM for L-pipecolic acid and sarcosine respectively; the V_{max} values were 16.9 and 14.92 nmol/min per mg of protein. These values were nearly identical with the K_m values of 1.9 and 6.7 mM and the V_{max} values of 9.8 and 12.3 nmol/min per mg of protein reported previously for a similar construct with the rabbit sarcosine oxidase.

Expression of L-pipecolic acid oxidase in mammalian cells reveals that it is targeted to peroxisomes by a PTS 1 mechanism

Human fibroblasts show no evidence of the presence of L-pipecolic acid oxidase either by enzyme activity or by indirect immunofluorescence studies with the affinity-purified polyclonal antibody against native L-pipecolic acid oxidase. However, when

normal human fibroblasts were transfected with the cDNA for L-pipecolic acid oxidase, the expressed protein was found in a punctate pattern that co-localized with the peroxisomal protein catalase (Figures 5A and 5B). In contrast, both L-pipecolic acid oxidase and catalase showed a diffuse pattern of cytosolic staining when L-pipecolic acid oxidase was transfected into fibroblasts from a patient with a specific defect in PTS 1 protein targeting (Figures 5C and 5D). These results support the hypothesis that this protein is imported into peroxisomes via the PTS 1 pathway.

Northern blot studies of L-pipecolic acid oxidase suggest that expression is mainly in liver and kidney

We probed a multiple-human-tissue Northern blot with the cDNA for L-pipecolic acid oxidase and found that a transcription product of 2.3 kb was expressed only in liver and kidney (Figure 6). Previous studies had suggested that L-pipecolic acid oxidase is active in the brain and that it might even be part of an important pathway for lysine degradation in the brain [24]. No signal was found in the composite sample from brain, even after prolonged exposure of the blot (results not shown).

Plasma L-pipecolic acid, but not sarcosine, levels are elevated in patients with PBDs

With the use of amino acid analysis, plasma sarcosine was undetectable in both a combined PBD patient plasma sample and in a control human plasma specimen. The same PBD patient specimen contained L-pipecolic acid levels of 112 μ M (94.6–125.9;

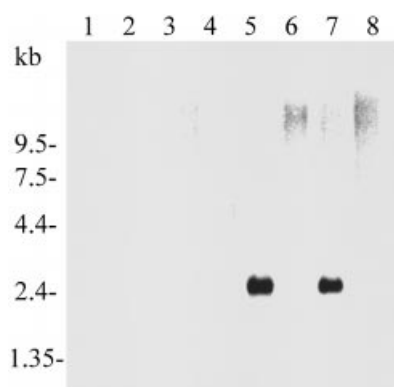


Figure 6 Northern blot of human L-pipecolic acid oxidase

The entire open reading frame of the gene was randomly labelled and the product was used to probe a human Northern blot as described in the Materials and methods section. Numbered lanes contained RNA from the following tissues: lane 1, heart; lane 2, brain; lane 3, placenta; lane 4, lung; lane 5, liver; lane 6, skeletal muscle; lane 7, kidney; lane 8, pancreas. The blot was processed with a Fuji PhosphorImager and was rendered with Adobe Photoshop 4.0.

$n = 6$), in comparison with control levels of $1.8 \mu\text{M}$ ($1.7\text{--}1.9$; $n = 6$).

Phylogenetic analysis of the sarcosine oxidase family suggests that the proteins separate into at least four major groups

We constructed an inferred phylogenetic tree (cladogram) (Figure 7) to analyse the evolutionary relationships between the proteins identified as actually or hypothetically involved in the degradation of sarcosine and related compounds (e.g. dimethylglycine and L-pipecolic acid). In all, 17 sarcosine-degrading proteins were identified in computer searches of public DNA and protein databases. Complete amino acid sequences were used for the monomeric proteins. For the heterodimers, only the subunit involved in the sarcosine oxidase reaction was analysed. Because the *C. elegans* gene included an apparent duplication, the two portions were treated as two different proteins. To construct the tree, the amino acid sequences of 18 sarcosine-degrading enzymes were multiply aligned and the character positions containing any gaps were eliminated. Parsimony analysis was used to construct a tree that used the minimum number of evolutionary changes to account for the differences at each amino acid position among the 18 sequences. The tree is unrooted because there is no ancestral sarcosine oxidase known to define an outgroup. In an unrooted tree such as Figure 7 there is no root node; branch lengths quantify relationships between the sarcosine-degrading proteins without defining a primordial evolutionary path.

A bootstrap analysis was performed to gain a statistical measure of confidence in the tree; 100 trees were generated from the initial data set, and the percentage of trees containing a particular clade was measured. (In this context a clade is a group of sarcosine-degrading enzymes containing a common ancestor that is not shared by any sarcosine-degrading enzyme outside the group.)

Four major clades were identified. L-Pipecolic acid oxidase is part of a clade consisting of five sarcosine oxidase sequences, comprising four complete proteins. All complete proteins in this clade terminate in presumed peroxisomal C-terminal PTS 1 targeting signals. At least two members of this family (the rabbit and human proteins) have also been verified as containing a covalently bound FAD. Members of the second clade, consisting

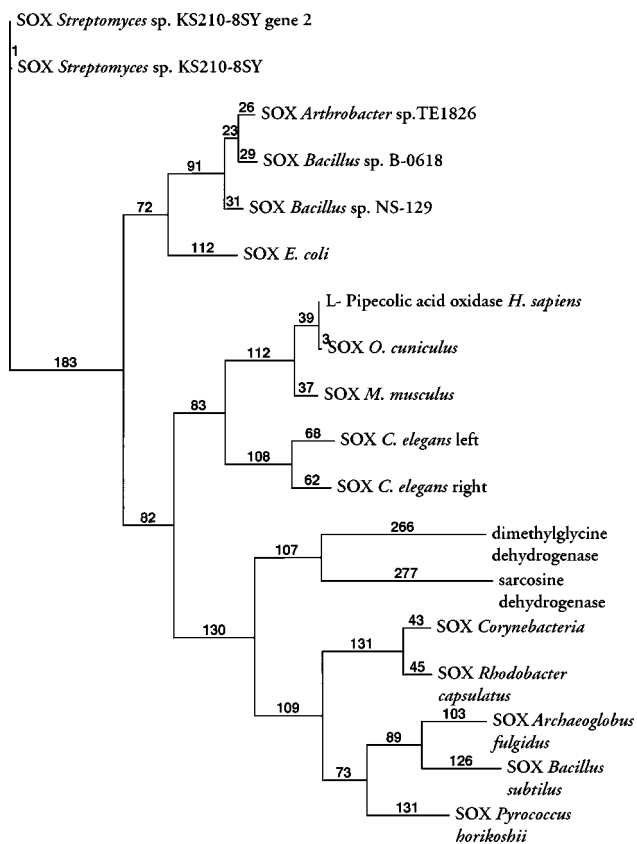


Figure 7 Phylogenetic analysis of the sarcosine oxidase (SOX) family of proteins

Sixteen full-length sarcosine/L-pipecolic acid oxidases/dehydrogenases were identified in the GenBank DNA database; the sequence for L-pipecolic acid oxidase was obtained from this study. Because the sequence from *C. elegans* suggested a gene duplication, the two portions were treated as two separate proteins. The tree represents an unrooted cladogram. Branch length values are indicated and are additive. The full names of the abbreviated species names given in the figure are *Homo sapiens*, *Oryctolagus cuniculus*, *Mus musculus*, *Caenorhabditis elegans* and *Escherichia coli*. Both the sarcosine dehydrogenase and dimethylglycine dehydrogenase sequences were derived from *Rattus norvegicus*.

of bacterial proteins such as the *Bacillus* sp. NS-129 products, have all been characterized as monomeric sarcosine oxidases. All members of the family that have been examined also contain a covalently bound FAD. The translational products of the two genes from *Streptomyces* sp. KB210-8SY form a third clade. These products are uncharacterized. A fourth clade consists of other bacterial sarcosine oxidases including the heterodimeric sarcosine oxidases such as those found in sarcosine dehydrogenase from *Corynebacteria* and *Archaeoglobus fulgidus*, as well as in two mammalian sarcosine dehydrogenases. In contrast with the monomeric sarcosine oxidases, the heteromeric *Corynebacteria* protein as well as the dehydrogenases in this clade transfer their methyl products to folate [25]. The two mammalian dehydrogenases both reside in mitochondria and are linked into the electron transport chain.

DISCUSSION

Almost as soon as PBDs were identified as a disease entity, the accumulation of L-pipecolic acid was recognized as diagnostic for these diseases [3]. Subsequently it was shown that L-pipecolic acid oxidase, the enzyme involved in its degradation in humans,

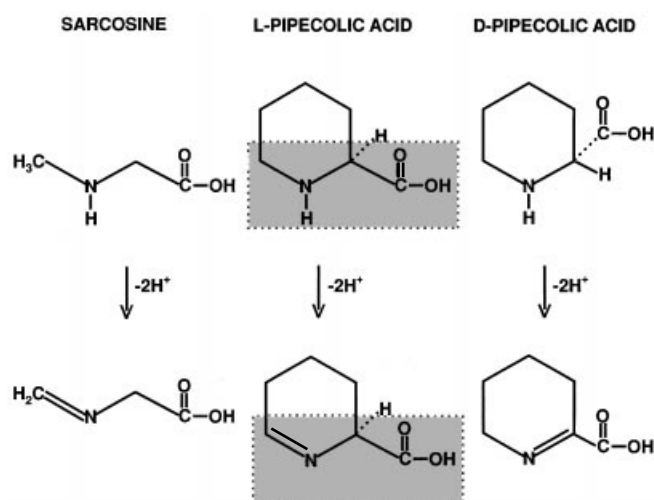


Figure 8 Comparison of reactions involved in the dehydrogenation of L-pipecolic acid, sarcosine and D-pipecolic acid

The dehydrogenation of L-pipecolic acid involves the formation of a double bond between C-6 and the nitrogen in the ring, whereas the dehydrogenation of sarcosine occurs between the same carbon and nitrogen (see the boxed area of L-pipecolic acid for comparison). In contrast, D-pipecolic acid is dehydrogenated between C-1 and nitrogen.

is located in peroxisomes [26] and is defective in PBDs [1]. On cloning the associated gene, we have found that its amino acid sequence shows a high degree of similarity to the sarcosine oxidases and that its protein product is capable of oxidizing both sarcosine and L-pipecolic acid. Surprisingly, no similarity was found to the peroxisomal D-amino acid oxidases. Furthermore, the protein product is targeted to peroxisomes, suggesting that its C-terminal AHL is a functional targeting signal.

Several lines of evidence suggest that the gene product identified by the reverse genetics approach is the same protein as the purified native L-pipecolic acid oxidase obtained from monkey liver. First, the translation of the gene yields sequences that are nearly identical with those obtained from the internal peptides of the purified protein. Secondly, the purified recombinant protein oxidizes L-pipecolic acid with kinetics similar to that of the purified native protein [5]. Thirdly, the chimaeric protein product reacts with the antibody generated against the original protein purified from monkey liver. Lastly, like the original protein, the protein product of this gene is located in peroxisomes.

The finding that this enzyme is most closely related to the sarcosine oxidases was unexpected. Originally we had hypothesized that the enzyme might be similar to the D-amino acid oxidases. However, on closer inspection, it is apparent that the reactions of L-pipecolic acid oxidase and that of monomeric sarcosine oxidase are essentially identical (Figure 8). When L-pipecolic acid, a six-membered-ring compound containing an imino group, is dehydrogenated to Δ^1 -piperidine-6-carboxylic acid, this reaction (Figure 8, boxed area) replicates the dehydrogenation of sarcosine. In contrast, D-amino acid oxidase forms a double bond on the opposite side of the nitrogen in D-pipecolic acid, yielding instead Δ^1 -piperidine-2-carboxylic acid [27]. Similarly, the fructosyl amino acid oxidases that were identified as somewhat similar to both groups of enzymes also oxidize a ketoamine to an aldehyde and an amine.

The first suggestion that L-pipecolic acid oxidase might be different from the D-amino acid oxidases came with the discovery that the protein contains a covalently bound flavin [6]. D-Amino

acid oxidases, as well as most other flavoproteins, use non-covalently bound flavins that are tightly associated with the protein. In contrast, all identified sarcosine oxidases and dehydrogenases, as well as fructosyl amino acid oxidases, contain at least one covalently bound flavin, either FAD or FMN. Willie et al. [28] identified the covalently bound flavin in one monomeric sarcosine oxidase as FAD, a finding identical with that for both L-pipecolic acid oxidase [6] and fructosyl amino acid oxidase [29].

The only sarcosine-degradating protein in which the exact site of FAD binding has been identified is dimethylglycine dehydrogenase [30]. We found by BLAST analysis that L-pipecolic acid oxidase and both sarcosine dehydrogenase and dimethylglycine dehydrogenase are similar around the N-terminal nucleotide-binding fold (region 1). With dimethylglycine dehydrogenase, this region of similarity extends to the histidine residue that has been identified as covalently binding FAD. Thus, by extension, in L-pipecolic acid oxidase the histidine residue at position 44 would seem to be a good candidate for the covalent FAD-binding site. However, during our amino acid sequencing of internal peptide 3, this amino acid appeared as a normal histidine residue, as also did His-33 and His-39. Furthermore, this peptide did not fluoresce when the excitation and emission filters appropriate for flavins were used, whereas at least two other peptides were identified that fluoresced. Thus the covalent flavin-binding site in L-pipecolic acid oxidase is probably not homologous with that reported for dimethylglycine dehydrogenase.

Particularly striking is the finding that so much amino acid sequence is conserved between certain regions of the bacterial monomeric sarcosine oxidases, human L-pipecolic acid oxidase, and certain regions of the fructosyl amino acid oxidases. All of these groups share an N-terminal $\beta\alpha\beta$ binding fold that is also common to dimethylglycine dehydrogenase from rat liver, to the β subunit of the heterotetrameric sarcosine oxidases from the *Corynebacterium* sp. P1 [31] and to amino acid deaminase from *Proteus mirabilis*. However, homologous regions 3 and 4 are uniquely conserved among these proteins. In these regions, the similarity between the monomeric sarcosine oxidase from *Bacillus* sp. NS-129 and human L-pipecolic acid oxidase is particularly great. In fact, 16 of 19 amino acids are identical and 18 of 19 are identical or conserved in the third site, whereas 13 of 17 amino acids are identical and 15 of 17 are identical or conserved in the fourth region. In contrast, the third region is not conserved in either the multimeric *Corynebacterium* sp. P1 sarcosine oxidase or dimethylglycine dehydrogenase [31], and the fourth is only mildly conserved (less than 30% identity).

When the fructosyl amino acid oxidases were compared with both L-pipecolic acid oxidase and sarcosine oxidase from *Bacillus* sp. NS-129, only 4 amino acids were identical in both regions 3 and 4 (of 19 and 17 amino acids respectively), but 9 of 19 and 6 of 17 amino acids were similar (Figure 3 and 4). These proteins cannot use sarcosine as a substrate; it is consistent with these findings that their amino acid sequences did not fit into the phylogenetic analysis as described. The finding of specific conservation in the third and fourth homologous regions among these diverse families of proteins suggests that these areas might be functionally important. Region 3 has already been shown to include the conserved cysteine residue (position 339) (Figure 4, bold type) that was predicted by site-directed mutagenesis to be the active site for sarcosine oxidase [32].

This conserved region might also house the covalent binding site for FAD. Homologous region 3 contains one histidine residue (position 354) that is absolutely conserved (Figure 4) and is therefore a candidate for this function. The only other histidine residue in this region (position 369) is conserved among

L-pipecolic acid oxidase and all of the monomeric sarcosine oxidases but is not conserved in the fructosyl amino acid oxidases. Future studies will address whether either of these histidine residues functions as the FAD-binding site.

Because this newly identified enzyme utilizes both L-pipecolic acid and sarcosine as substrates, the obvious next question is whether this enzyme is functionally a sarcosine oxidase or an L-pipecolic acid oxidase. Kinetic studies comparing catalytic constants for the rabbit protein suggested that L-pipecolic acid is only slightly more favoured as a substrate [14]. The kinetic values for the human protein are essentially identical. However, the finding that sarcosine levels are normal in patients with PBDs is strong evidence that L-pipecolic acid is the physiologically important substrate for this enzyme in humans. In contrast, when mitochondrial sarcosine dehydrogenase is blocked in glutaric aciduria type II, the levels of both L-pipecolic acid and sarcosine are elevated (R. I. Kelley, personal communication). It is unclear at present whether this accumulation occurs because the common transport system for both imino acids is overwhelmed or because there is insufficient L-pipecolic acid oxidase to process both substrates.

The oxidation of L-pipecolic acid has proved to be more complex than was originally expected. Earlier studies showed that L-pipecolic acid [26] is oxidized in different organelles in different species of animals. For example, in rabbits, 80% of the degradation activity resides in mitochondria (G. Dodt, unpublished work) and this activity uses a classical dehydrogenase mechanism [5]. Dogs, sheep and guinea pigs all exhibit this mitochondrial activity. Another 20% of the potential activity in rabbits is found in peroxisomes [14], whereas this peroxisomal activity is the only L-pipecolic acid degradation activity identified in humans and primates. In addition, the oxidation of L-pipecolic acid in rabbit mitochondria was associated with the electron transport chain and required exogenous FAD. In contrast, the peroxisomal enzyme activity in human and monkey tissues was inhibited by artificial electron acceptors and uninfluenced by exogenous FAD.

Because the peroxisomal enzyme shows similarity to the sarcosine oxidases and dehydrogenases, one might postulate that sarcosine dehydrogenase is a likely candidate for the mitochondrial L-pipecolic acid-degrading enzyme. However, the dependence of the activity on exogenous FAD rules out sarcosine dehydrogenase, which contains a covalently linked flavin [30]. A more probable candidate for this activity is the mitochondrial proline dehydrogenase. With L-pipecolic acid's structural similarity to the five-carbon-ring imino acid proline, one might expect its degradative enzyme to be similar. In fact, both the purified L-pipecolic acid oxidase from monkeys and sarcosine oxidase from rabbit kidney displayed some ability to oxidize L-proline [10]. Although we have no idea what enzyme is involved in mitochondria, its dependence on exogenous FAD suggests that it probably evolved independently of the sarcosine-degrading enzymes.

Nevertheless, in both monkeys and humans, L-pipecolic acid is degraded by an oxidase that occurs only in peroxisomes [1,5] and this protein is most similar to the monomeric sarcosine oxidases. Moreover, its accumulation when peroxisomes fail to assemble normally verifies that, in humans, peroxisomes are the only site for the degradation of L-pipecolic acid. Apparently, enzymes similar to L-pipecolic acid oxidase are found in both rabbits and

C. elegans. What is not clear at present is whether enzymes found in other branches of this protein family can oxidize L-pipecolic acid or whether this branch is unique in its ability to utilize L-pipecolic acid as a substrate.

We thank Dr. P. A. Watkins, Dr. R. I. Kelley and Dr. M. S. Jorns for their thoughtful advice, Dr. H. W. Moser for his generous advice and support, Dr. G. Thomas for performing the sarcosine quantitation on plasma, and Dr. J. Pevsner for his help in compiling the phylogenetic tree. This work was supported by NIH grants HD10981 and HD24081.

REFERENCES

- Mihalik, S. J., Moser, H. W., Watkins, P. A., Danks, D. M., Poulos, A. and Rhead, W. J. (1989) *Pediatr. Res.* **25**, 548–552
- Wanders, R. J. A., Romeyn, G. J., vanRoermond, C. W. T., Schutgens, R. B. H., vandenBosch, H. and Tager, J. M. (1988) *Biochem. Biophys. Res. Commun.* **154**, 33–38
- Thomas, G., Haslam, R. H. A., Batshaw, M. L., Capute, A. J., Neidengard, L. and Ransom, J. L. (1975) *Clin. Genet.* **8**, 376–382
- Basso, L. V., Rao, R. D. and Rodwell, V. W. (1962) *J. Biol. Chem.* **237**, 2239–2245
- Mihalik, S. J. and Rhead, W. J. (1989) *J. Biol. Chem.* **264**, 2509–2517
- Mihalik, S. J. and McGuinness, M. C. (1991) in *Flavins and Flavoproteins 1990* (Curti, B., Ronchi, S. and Zanetti, G., eds.), pp. 881–884, Walter de Gruyter, New York
- Lazarow, P. B. and Moser, H. W. (1995) in *The Metabolic and Molecular Bases of Inherited Disease* (Scriver, C. R., Beaudet, A. L., Sly, W. S. and Valle, D., eds.), pp. 2287–2324, McGraw-Hill, New York
- Chang, Y.-F. and Myslinski, N. R. (1985) *Neurosci. Lett.* **59**, 79–84
- Gutierrez, M. D. C. and Giacobini, E. (1985) *Neurochem. Res.* **10**, 691–702
- Mihalik, S. J., McGuinness, M. and Watkins, P. A. (1991) *J. Biol. Chem.* **266**, 4822–4830
- Matsudaira, P. T. (1989) *A Practical Guide to Protein and Peptide Purification for Microsequencing*, Academic Press, San Diego
- Lee, C. C. and Caskey, C. T. (1990) in *PCR Protocols* (Innis, M. A., Gelfand, D. H., Sninsky, J. J. and White, T. J., eds.), pp. 46–53, Academic Press, San Diego
- Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd edn., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
- Reuber, B. E., Karl, C., Reimann, S. A., Mihalik, S. J. and Dodt, G. (1997) *J. Biol. Chem.* **272**, 6766–6776
- Laemmli, U. K. (1970) *Nature (London)* **227**, 680–685
- O'Farrell, P. H. (1975) *J. Biol. Chem.* **250**, 4007–4021
- Thompson, J. D., Higgins, D. G. and Gibson, T. J. (1994) *Nucleic Acids Res.* **22**, 4673–4680
- Slawecki, M. L., Dodt, G., Steinberg, S., Moser, A. B., Moser, H. W. and Gould, S. J. (1995) *J. Cell Sci.* **108**, 1817–1829
- Bradford, M. M. (1976) *Anal. Biochem.* **72**, 248–254
- Gould, S. J., Keller, G. A., Hosken, N., Wilkinson, J. and Subramani, S. (1989) *J. Cell Biol.* **108**, 1657–1664
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990) *J. Mol. Biol.* **215**, 403–410
- Herbst, R., Barton, J. L. and Nicklin, M. J. (1997) *Genomics* **46**, 480–482
- Wierenga, R. K., Terpstra, P. and Hol, W. G. (1986) *J. Mol. Biol.* **187**, 101–107
- Rao, V. V., Tsai, M. J., Pan, X. M. and Chang, Y. F. (1993) *Biochim. Biophys. Acta* **1164**, 29–35
- Wagner, M. A. and Jorns, M. S. (1997) *Arch. Biochem. Biophys.* **342**, 176–181
- Mihalik, S. J. and Rhead, W. J. (1991) *J. Comp. Physiol. B* **160**, 671–675
- Chang, Y. F. and Adams, E. (1974) *J. Bacteriol.* **117**, 753–764
- Willie, A., Edmonson, D. E. and Jorns, M. S. (1996) *Biochemistry* **35**, 5292–5299
- Yoshida, N., Sakai, Y., Serata, M., Tani, Y. and Kato, N. (1995) *Appl. Environ. Microbiol.* **61**, 4487–4489
- Cook, R. J., Misono, K. S. and Wagner, C. (1985) *J. Biol. Chem.* **260**, 12998–13002
- Chlumsky, L. J., Zhang, L. and Jorns, M. S. (1995) *J. Biol. Chem.* **270**, 18252–18259
- Nishiya, Y., Zuihara, S. and Imanaka, T. (1995) *Appl. Environ. Microbiol.* **61**, 367–370