Prostaglandins, Leukotrienes and Essential Fatty Acids (1999) 60(5&6), 323–328

Article no. plef.1999.0045

Human very long-chain acyl-CoA synthetase and two human homologs: initial characterization and relationship to fatty acid transport protein

P. A. Watkins, J. Pevsner, S. J. Steinberg

Kennedy Krieger Institute and Departments of Neurology and Neuroscience, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA

Summary Several human genes with a high degree of homology to rat very long-chain acyl-CoA synthetase (rVLCS) and mouse fatty acid transport protein (mFATP) were identified. Full-length cDNA clones were obtained for three genes, and predicted amino acid sequences were generated. Initial characterization indicated that one gene was most likely hVLCS, the human ortholog of rVLCS. The other two (hVLCS-H1 and hVLCS-H2) were more closely related to rVLCS than to mFATP. Phylogenetic analysis of amino acid sequences confirmed that hVLCS-H1 and hVLCS-H2 were evolutionarily closer to VLCSs than FATPs. Alignment of predicted amino acid sequences of human, rat and mouse VLCSs and FATPs revealed the existence of two highly conserved motifs. While one motif is also present in long-chain acyl-CoA synthetases, the other serves to distinguish the VLCS/FATP family from the long-chain synthetase family. Elucidation of the biochemical functions of all VLCS/FATP family members should provide new insights into cellular fatty acid metabolism.

INTRODUCTION

Fatty acids are activated to their coenzyme A (CoA) thioesters for entry into the metabolic stream.[1] Enzymes catalyzing fatty acid activation have been classified by their chain length specificities into short-, medium-, long- and very long-chain acyl-CoA synthetases.[1] Short-chain acyl-CoA synthetases activate acetate, propionate and butyrate, and medium-chain enzymes activate fatty acids containing 6–10 carbon.[2–4] Long-chain acyl-CoA synthetases (LCS) are capable of activating the primary fatty acids of dietary origin, palmitic acid and oleic acid, and are the best characterized of this group. However, the optimal substrate chain length for LCS is 12–14 carbons.[5,6] Very long-chain fatty acids (VLCFA) containing 22 or more carbons are also activated in cells and tissues.[7–10] Because LCS is relatively inactive with these substrates, the existence of very long-chain acyl-CoA synthetase (VLCS) was postulated.

VLCS activity is found in microsomal and peroxisomal fractions of liver, brain and other tissues; no VLCS activity is associated with mitochondria.[7,8] In contrast, LCS activity is present in all three organelles.[6] Rat VLCS (rVLCS) was purified from liver peroxisomes in 1996.[11] While its chain length specificity was not studied in detail, the purified enzyme activated the long-chain fatty acid, palmitate (C16:0) at a rate 1.5-fold higher than lignoceric acid (C24:0), a VLCFA. Immunoblot analysis of subcellular fractions with antibody raised against purified VLCS verified that the protein was not detectable in mitochondria.[12] Immunoreactive VLCS was found in microsomes as well as peroxisomes, in agreement with enzyme activity measurements. Shortly after its purification, rVLCS cDNA was cloned and sequenced.[12] Comparison of the predicted amino acid sequence of rVLCS to other known proteins revealed that it was most closely related to fatty acid transport protein (FATP), a plasma membrane protein.[13]

Correspondence to: P. A. Watkins, Kennedy Kreger Institute, 707 N. Broadway, Baltimore, MD 21205, USA. Tel.: +1 410 502 9493; Fax: +1 410 502 8279; E-mail: watkins@kennedykrieger.org

This work was supported by NIH grants HD10981, NS10533 and HD24061.

Homology of rVLCS to the three known rat LCSs was weaker.

The biological importance of VLCS is underscored by the existence of human diseases in which VLCFA accumulate. In both the peroxisome biogenesis disorders (PBD, which include the clinical phenotypes Zellweger syndrome, neonatal adrenoleukodystrophy, and infantile Refsum disease) and X-linked adrenoleukodystrophy (X-ALD), VLCFA accumulate as a result of a decreased rate of catabolism.[14,15] Examination of cells from PBD and X-ALD patients revealed that VLCFA β-oxidation takes place in peroxisomes and not mitochondria.[16] This finding is in concert with the observation that VLCS activity is associated with peroxisomes but not mitochondria.[7,8] While failure of peroxisomes to form normally is the underlying cause of VLCFA accumulation in PBD patients, X-ALD patients were found to have decreased peroxisomal VLCS activity.[9,10] Therefore, it was of interest to characterize the human VLCS in order to investigate the biochemical defect in X-ALD. We report here that in addition to the human ortholog (hVLCS) of rVLCS, humans contain at least five related genes. We have cloned cDNAs encoding hVLCS and two homologs, and have begun to characterize their properties.

EXPERIMENTAL PROCEDURES

Identification of human homologs of rVLCS

To search for human proteins homologous to rVLCS, the BLAST algorithm[17] was used to probe the NCBI expressed seauence tag (EST) database. Initially, the query sequence was either full or partial rVLCS amino acid sequence. Human ESTs with homology to rVLCS were identified and these EST sequences were used as query in a second round of BLAST searching. A total of 72 ESTs were identified by this method. Overlapping EST sequences were identified, aligned and assembled into contigs using the Sequencher program (ver. 2.1, Gene Codes Corp., Inc.).

Cloning of three human VLCS homologs

Full-length cDNA clones of three human genes encoding proteins homologous to rVLCS were obtained. Details of the cloning procedures will be published elsewhere (S. J. Steinberg et al., manuscripts in preparation). In brief, the starting point for each cloning was the EST clone containing the longest open reading frame. These ESTs generally contained the stop codon and some 3′-untranslated region. To obtain the remaining unknown 5′-sequence, one or more additional techniques were employed, including polymerase chain reaction, phage library screening and 5′-rapid amplification of cDNA ends.

Sequence analysis

Multiple alignment of amino acid sequences and calculation of percent amino acid identity and similarity were done using the ClustalW program[18] or the pileup program of the GCG software package (Madison, WI). Hydropathy analysis was done using the algorithm of Kyte and Doolittle[19] (window size = 19 amino acids) using the DNA Strider program (ver. 1.2). Sequences were analyzed for signal sequences using the SignalP prediction server[20] (url: http://www.cbs.dtu.dk/services/SignalP/). The phylogenetic tree was generated using the Phylogenetic Analysis Using Parsimony program (PAUP version 4.0.0d64, kindly provided by Dr David Swofford, Smithsonian Institution).

RESULTS

Human homologs of rVLCS

Homology probing of the human EST database revealed the existence of several genes that were homologous to rVLCS. Over 70 different human ESTs were identified that were subsequently assembled into six distinct contigs. These contigs varied in length, and none were considered to be long enough to contain complete open reading frames. Comparison of the predicted amino acid sequences of each contig to the corresponding region of rVLCS indicated that one had a very high degree of homology and, therefore, was most likely the human ortholog, hVLCS. Comparison of predicted amino acid sequences to that of mouse FATP (mFATP) indicated that another of the contigs likely represented hFATP. Of the remaining four proteins predicted from contigs, three were more homologous to VLCS, while one was more homologous to FATP.

Cloning and characteristics of hVLCS and two homologs, hVLCS-H1 and hVLCS-H2

In order to investigate the biochemical properties of hVLCS and its homologs, we obtained full-length cDNA clones encoding three distinct proteins. rVLCS cDNA contains an open reading frame encoding 620 amino acids[12] and mFATP cDNA encodes a protein of 646 amino acids.[13] Therefore, we expected that the corresponding human proteins would be of similar length. Full-length cDNA encoding the human protein predicted to be the ortholog of rVLCS was cloned. This cDNA also had an open reading frame encoding 620 amino acids. As shown in Table 1, hVLCS and rVLCS share 82.4% amino acid identity and 93.7% similarity when conservative substitutions are considered. In contrast, hVLCS and mFATP are only 41.0% identical and 57.7% similar.

We then determined which contig represented the

Table 1 Amino acid sequence homology of hVLCS, hVLCS-H1, hVLCS-H2 and related mammalian proteins

	hVLCS	hVLCS-H1	hVLCS-H2
rVLCS	82.4/93.7	48.8/73.2	45.0/68.5
hVLCS-H1	50.9/74.0	—	43.7/67.2
hVLCS-H2	45.3/69.4	43.7/67.2	—
mVLCSR	45.0/70.2	42.5/68.8	71.4/87.1
mFATP	41.0/57.7	38.4/61.7	39.9/58.7

The predicted amino acid sequences of hVLCS, hVLCS-H1, hVLCS-H2, rVLCS, mVLCSR and mFATP were aligned using the ClustalW multiple sequence alignment program[18] and the % identity (left of slash) and % similarity (right of slash) were calculated.

human gene next most homologous to rVLCS. Its full-length cDNA was cloned and designated hVLCS-homolog 1 (hVLCS-H1). hVLCS-H1 cDNA contains an open reading frame encoding 619 amino acids. This homolog is 50.9 % identical and 74.0% similar to hVLCS at the amino acid level (Table 1). Its homology to rVLCS is only slightly lower, with 48.8% identity and 73.2% similarity (Table 1). The homology of hVLCS-H1 to mFATP was lower, with 38.4% amino acid identity and 61.7% similarity.

A third human cDNA homologous to rVLCS, designated hVLCS-homolog 2 (hVLCS-H2), was also cloned. The open reading frame was somewhat longer, encoding 690 amino acids. The predicted amino acid sequence was found to be 45.0–45.3% identical and 68.5–69.4% similar to rVLCS and hVLCS (over 620 amino acid residues; Table 1). The degree of homology of hVLCS-H2 to hVLCS-H1 was not significantly different, with 43.7% identity and 67.2% similarity. Again, the homology to mFATP was weaker (39.9% identity and 58.7% similarity; Table 1). hVLCS-H2 is very closely related to the recently described 689 amino acid mVLCSR (mouse VLCS-related protein);[21] these proteins are 71.4% identical and 87.1% similar (Table 1).

To further probe the relationship between the proteins encoded by the three cloned human cDNAs and known rat and mouse genes, a phylogenetic tree was generated. Eight full-length protein sequences were multiply aligned and converted to a phylogenetic tree using parsimony analysis. An exhaustive search evaluated all 10 395 possible trees constructed from the 8 taxa and revealed a tree with the shortest branch length (Fig. 1). This tree, depicted as an unrooted phylogram, shows the relationships among the eight proteins consistent with the individual protein comparisons of Table 1. The branch lengths from hVLCS-H1 are closer to the VLCS than to hVLCS-H2 or mFATP. But since hVLCS-H1 is distinct from the three VLCS proteins, the hVLCS-H1 gene may be paralogous (derived from an ancestral duplication event) rather than orthologous (derived from a speciation event) to the VLCS group. Furthermore, hVLCS-H2 and its likely ortholog, mVLCSR, are more related to the VLCS than to the FATPs.

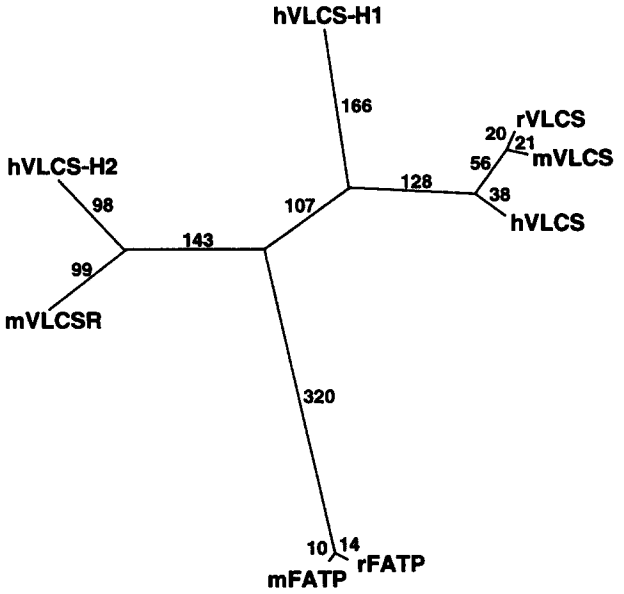

Fig. 1 Phylogenetic tree of VLCS/FATP proteins. Phylogenetic relationships were estimated from an alignment of the amino acid sequences. An exhaustive search using the PAUP program (see Experimental Procedures) revealed the single most parsimonious tree with the shortest total branch lengths. The branch lengths reflect inferred evolutionary distance and are related to the number of amino acid changes in each branch.

Potential targeting signals in hVLCS, hVLCS-H1 and hVLCS-H2

Because VLCS activity is associated with both peroxisomes and the endoplasmic reticulum, amino acid sequences of hVLCS, hVLCS-H1 and hVLCS-H2 were examined for potential targeting signals. Both rVLCS and hVLCS end in the carboxy-terminal tripeptide -LKL, which is similar to the consensus peroxisome targeting signal 1 (PTS 1), -SKL.[22] The carboxy-terminus of hVLCS-H1 ends in a related sequence, -IKL. While these sequences are all similar, neither -LKL nor -IKL have previously been demonstrated to be functional variants of the consensus PTS 1 signal in mammals. The tripeptide -WRL, which is not likely to be a functional PTS 1, was found at the carboxy-terminus of hVLCS-H2. PTS 2, a targeting signal that resides near the amino terminus of a small number of peroxisome proteins,[23] was not found in any of the three human proteins. Hydropathy analysis using the algorithm of Kyte and Doolittle[19] suggested that all three proteins had potential signal sequences at their amino termini for targeting to the endoplasmic reticulum (Fig. 2). Analysis of these sequences using the SignalP prediction server[20] revealed that they fulfilled some but not all criteria for signal sequences.

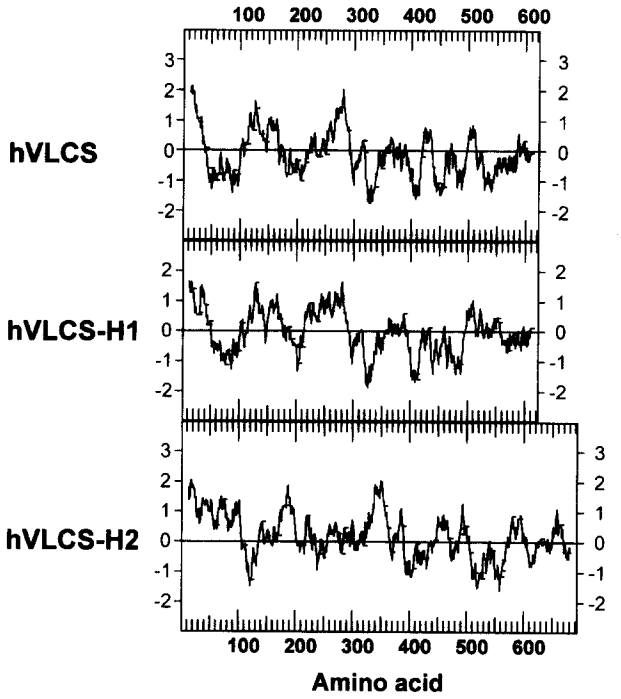

Fig. 2 Hydropathy analysis of human VLCS homologs. The predicted amino acid sequences of hVLCS, hVLCS-H1 and hVLCS-H2 were analyzed using the algorithm of Kyte and Doolittle with a window size of 19. Hydrophobic regions of proteins have positive values on the abscissa, while hydrophilic regions have negative values.

Motifs conserved between the VLCS and FATP protein families

Alignment of the amino acid sequences of hVLCS, hVLCS-H1, hVLCS-H2 and other mammalian VLCSs and FATPs revealed two motifs that were very highly conserved (Fig. 3). The first is a 10 amino acid sequence corresponding to amino acids 224–233 of hVLCS. This motif is found not only in the VLCS/FATP family, but also in all LCSs and in several other enzymes whose reaction mechanisms proceed via adenylated intermediates.[24] The second motif consists of 22 residues and corresponds to amino acids 476–497 of hVLCS. In contrast to motif 1, motif 2 is unique to the VLCS/FATP family. Out of 22 amino acids, 17 are invariant in the mammalian VLCS/FATP family. The corresponding region of LCSs contains 23 amino acids. However, only five residues are present in human LCS, and only four of these (shown in bold in Fig. 3) are invariant among all known LCSs. Thus, motif 2 serves to distinguish the VLCS/FATP family from other acyl-CoA synthetases. Analysis of the VLCS/FATP proteins with the Identify software program (url:http://dna.Stanford.EDU/identify/) revealed an adenylate kinase signature present within motif 2 (GDX[FILMV]R, where X is any amino acid and [FILMV] is any of these five amino acid residues). While the VLCS/FATP proteins do not share slgnificant overall amino acid identity to the adenylate kinases, the presence of this limited homology is consistent with the hypothesis that motif 2 forms part of a catalytic site.

DISCUSSION

Despite nearly a century of biochemical investigation, many aspects of fatty acid metabolism are still incompletely understood. Many new enzymes and other proteins involved have been identified over the last few years. The discoveries of some, e.g. very long-chain acyl-CoA dehydrogenase[25] and the adrenoleukodystrophy protein,[26] were unexpected. Others, e.g. VLCS and FATP, were predicted to exist but purification and cloning were

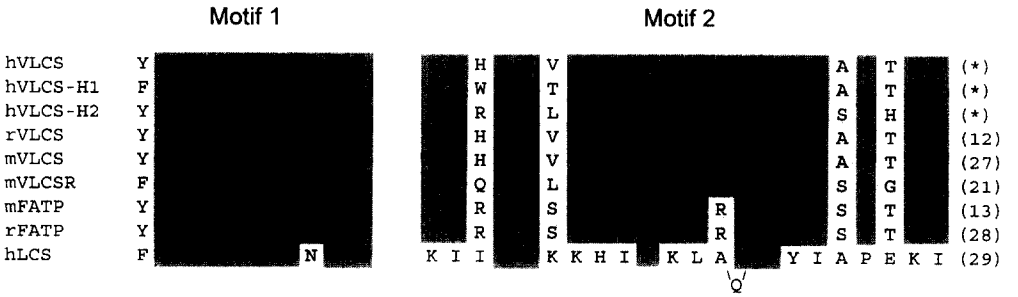

Fig. 3 Conserved motifs in the VLCS/FATP protein family. Two highly conserved motifs define the VLCS/FATP family and the second of these distinguishes this family from the LCS family. Motifs 1 and 2 correspond to amino acids 224–233 and 476–497, respectively, of hVLCS. Sequences of related proteins from human, mouse and rat are shown. These include the originally described VLCS from rat peroxisomes (rVLCS) and its orthologs hVLCS and mVLCS, the two human homologs of hVLCS described in this article (hVLCS-H1 and hVLCS-H2), the mouse VLCS-related protein (mVLCSR; orthologous to hVLCS-H2) and mouse and rat FATPs (mFATP and rFATP). The comparable region of a representative human LCS (hLCS; HUMLCACS) is also shown. Identical amino acid residues are shaded. Residues conserved between both the VLCS/FATP family and the LCS family are in bold typeface. Sources of sequences are shown on the right. *S. J. Steinberg et al., manuscripts submitted.

only recently achieved.[12,13] As sequencing of human and other genomes progresses, still more proteins involved in fatty acid metabolism are almost certain to be discovered. As reported here, we have cloned the human ortholog of rVLCS and two additional homologs. While it was presumed that hVLCS was present in human tissues, the existence of hVLCS-H1 and hVLCS-H2 was not predicted.

The biochemical functions of hVLCS, hVLCS-H1 and hVLCS-H2 are currently under investigation. Preliminary results indicate that like the rat enzyme, hVLCS is capable of activating both VLCFA and long-chain fatty acids (S. J. Steinberg et al., manuscript in preparation). Thus far, we have not been able to demonstrate unequivocally that either hVLCS-H1 or hVLCS-H2 have significant acyl-CoA synthetase activity. Since only a limited number of substrates have been examined, it would be premature to conclude that these homologs are not acyl-CoA synthetases. Moreover, hVLCS-H1 or hVLCS-H2 might require the presence of one or more other proteins, e.g. the adrenoleukodystrophy protein (ALDP), to exhibit VLCS activity. While the relationship between ALDP and VLCS activity has not been clearly defined, it has been shown that when ALDP is defective, peroxisomal VLCS activity is diminished[9,10] The prospect that either hVLCS-H1 or hVLCS-H2 requires ALDP for expression of VLCS activity is currently under investigation.

We must also consider the possibility that hVLCS-H1 and hVLCS-H2 have other functions in cellular fatty acid metabolism. One such function is to facilitate fatty acid transport into cells or across organelle membranes. Facilitating transport into cells is improbable because preliminary experiments suggest that, unlike FATP, neither protein is located in the plasma membrane (S. J. Steinberg et al., manuscript in preparation). Further characterization of these proteins based upon subcellular localization and tissue expression patterns may provide additional clues as to their biochemical functions.

Although the precise functions of hVLCS-H1 and -H2 have not yet been defined, it is almost certain that they are involved in fatty acid metabolism. They are closely related to human, rat and mouse VLCSs, and only slightly more distantly related to rat and mouse FATPs. They contain a motif (motif 1) common to all known acyl-CoA synthetases. Furthermore, they contain an amino acid sequence (motif 2) that we propose is the signature motif of the VLCS/FATP protein family. Motif 2 was also found in the predicted amino acid sequences of two of the three human EST contigs for which full-length cDNA has not yet been cloned (P. A. Watkins, unpublished observations); not enough information is available to determine whether this motif is in the third contig. The elucidation of the functions of all the human VLCS homologs should provide exciting new insights into cellular fatty acid metabolism.

ACKNOWLEDGEMENTS

The authors would like to thank Susan J. Wang, Do G. Kim, Sai Lui and Emily Chen for excellent technical assistance and Drs Kirby D. Smith, Lelita T. Braiterman, Stephan Kemp, Stephanie J. Mihalik and Hugo W. Moser for many stimulating discussions of this work.

REFERENCES

1. Watkins P. A. Fatty acid activation. *Prog Lipid Res* 1997; **36**: 55–83.
2. Webster L. T. Jr. Studies of the acetyl coenzyme A synthetase reaction. II. Crystalline acetyl coenzyme A synthetase. *J Biol Chem* 1965; **240**: 4158–4163.
3. Webster L. T. Jr., Gerowin L. D., Rakitak L. Purification and characteristics of a butyryl coenzyme A synthetase from bovine heart mitochondria. *J Biol Chem* 1965; **240**: 29–33.
4. Mahler H. R., Wakil S. J., Bock R. M. Studies on fatty acid oxidation. I. Enzymatic activation of fatty acids. *J Biol Chem* 1953; **204**: 453–468.
5. Kornberg A., Pricer W. E. Jr. Enzymatic synthesis of the coenzyme A derivatives of long chain fatty acids. *J Biol Chem* 1953; **204**: 329–343.
6. Tanaka T., Hosaka K., Hoshimaru M., Numa S. Purification and properties of long-chain acyl-coenzyme-A synthetase from rat liver. *Eur J Biochem* 1979; **98**: 165–172.
7. Singh H., Poulos A. Distinct long chain and very long chain fatty acyl CoA synthetases in rat liver peroxisomes and microsomes. *Arch Biochem Biophys* 1988; **266**: 486–495.
8. Singh I., Lazo O., Kremser K. Purification of peroxisomes and subcellular distribution of enzyme activities for activation and oxidation of very-long-chain fatty acids in rat brain. *Biochim Biophys Acta* 1993; **1170**: 44–52.
9. Lazo O., Contreras M., Hashmi M., Stanley W., Irazu C., Singh I. Peroxisomal lignoceroyl-CoA ligase deficiency in childhood adrenoleukodystrophy and adrenomyeloneuropathy. *Proc Natl Acad Sci USA* 1988; **85**: 7647–7651.
10. Wanders R. J. A., van Roermund C. W. T., van Wijland M. J. A. et al. Direct demonstration that the deficient oxidation of very long chain fatty acids in X-linked adrenoleukodystrophy is due to an impaired ability of peroxisomes to activate very long chain fatty acids. *Biochem Biophys Res Commun* 1988; **153**: 168–624.
11. Uchida Y., Kondo N., Orii T., Hashimoto T. Purification and properties of rat liver peroxisomal very-long-chain acyl-CoA synthetase. *J Biochem (Tokyo)* 1996; **119**: 565–571.
12. Uchiyama A., Aoyama T., Kamijo K. et al. Molecular cloning of cDNA encoding rat very long-chain acyl-CoA synthetase. *J Biol Chem* 1996; **271**: 30360–30365.
13. Schaffer J. E., Lodish H. F. Expression cloning and characterization of a novel adipocyte long chain fatty acid transport protein. *Cell* 1994; **79**: 427–436.
14. Lazarow P B., Moser H. W. Disorders of peroxisome biogenesis. In: Scriver C. R., Beaudet A. L., Sly W. S., Valle D., (Eds). The Metabolic and Molecular Bases of Inherited Disease. New York: McGraw-Hill, 1995: 2287–2324.
15. Moser H. W., Smith K. D., Moser A. B. X-linked adrenoleukodystrophy. In: Scriver C. R., Beaudet A. L., Sly W. S., Valle D., (Eds). The Metabolic and Molecular Bases of Inherited Disease. New York: McGraw-Hill, 1995: 2325–2349.
16. Singh I., Moser A. B., Goldfischer S., Moser H. W. Lignoceric acid is oxidized in the peroxisome: implications for the Zellweger cerebro-hepato-renal syndrome and adrenoleukodystrophy. *Proc Natl Acad Sci USA* 1984; **81**: 4203–4207.

17. Altschul S. F., Madden T. L., Schaffer A. A. et al. Gapped BLAST and Psi-BLAST: a new generation of protein database search programs. *Nucl Acids Res* 1997; **25**: 3389–3402.
18. Thompson J. D., Higgins D. G., Gibson T. J. Clustal W. Improving the sensitivity of progressive multiple sequence alignments through sequence weighing, position specific gap penalties and weigh matrix choice. *Nucl Acids Res* 1994; **22**: 4673–4680.
19. Kyte J., Doolittle R. F. A simple method for displaying the hydrophobic character of a protein. *J Mol Biol* 1982; **157**: 105–132.
20. Nielsen H., Engelbrecht J., Brunak S., von Heijne G. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Prot Engin* 1997; **10**: 1–6.
21. Berger J., Truppe C., Neumann H., Forss-Petter S. A novel relative of the very-long-chain acyl-CoA synthetase and fatty acid transporter protein genes with a distinct expression pattern. *Biochem Biophys Res Commun* 1998; **247**: 255–260.
22. Gould S. J., Keller G. A., Hosken N., Wilkinson J., Subramani S. A conserved tripeptide sorts proteins to peroxisomes. *J Cell Biol* 1989; **108**: 1657–1664.
23. Subramani S. Protein import into peroxisomes and biogenesis of the organelle. *Ann Rev Cell Biol* 1993; **9**: 445–478.
24. Babbitt P. C., Kenyon G. L., Martin B. M. et al. Ancestry of the 4-chlorobenzoate dehalogenase – Analysis of amino acid sequence identities among families of acyl-adenyl ligases, enoyl-CoA hydratases/isomerases, and acyl-CoA thioesterases. *Biochemistry* 1992; **31**: 5594–5604.
25. Izai K., Uchida Y., Orii T., Yamamoto S., Hashimoto T. Novel fatty acid beta-oxidation enzymes in rat liver mitochondria. 1. Purification and properties of very-long-chain acyl-coenzyme-A dehydrogenase. *J Biol Chem* 1992; **267**: 1027–1033.
26. Mosser J., Douar A. M., Sarde C. O. et al. Putative X-linked adrenoleukodystrophy gene shares unexpected homology with ABC transporters. *Nature* 1993; **361**: 726–730.
27. Berger J., Truppe C., Neumann H., Forss-Petter S. cDNA cloning and mRNA distribution of a mouse very long-chain acyl-CoA synthetase. *FEBS Lett* 1998; **415**: 305–309.
28. Schaap F. G., Hamers L., Von der Vusse G. J., Glatz J. F. C. Molecular cloning of fatty acid-transport protein cDNA from rat. *Biochim Biophys Acta* 1997; **1354**: 29–34.
29. Abe T., Fujino T., Fukuyama R. et al. Human long-chain acyl-CoA synthetase – Structure and chromosomal location. *J Biochem* 1992; **111**: 123–128.

