

# Evidence for 26 distinct acyl-coenzyme A synthetase genes in the human genome<sup>S</sup>

Paul A. Watkins,<sup>1,\*†</sup> Dony Manguel,<sup>\*,†</sup> Zhenzhen Jia,<sup>\*</sup> and Jonathan Pevsner<sup>\*,§</sup>

Kennedy Krieger Institute\* and Departments of Neurology<sup>†</sup> and Neuroscience,<sup>§</sup> Johns Hopkins University School of Medicine, Baltimore, MD 21205

**Abstract** Acyl-coenzyme A synthetases (ACSs) catalyze the fundamental, initial reaction in fatty acid metabolism. “Activation” of fatty acids by thioesterification to CoA allows their participation in both anabolic and catabolic pathways. The availability of the sequenced human genome has facilitated the investigation of the number of ACS genes present. Using two conserved amino acid sequence motifs to probe human DNA databases, 26 ACS family genes/proteins were identified. ACS activity in either humans or rodents was demonstrated previously for 20 proteins, but 6 remain candidate ACSs. For two candidates, cDNA was cloned, protein was expressed in COS-1 cells, and ACS activity was detected. Amino acid sequence similarities were used to assign enzymes into subfamilies, and subfamily assignments were consistent with acyl chain length preference. Four of the 26 proteins did not fit into a subfamily, and bootstrap analysis of phylograms was consistent with evolutionary divergence. Three additional conserved amino acid sequence motifs were identified that likely have functional or structural roles. The existence of many ACSs suggests that each plays a unique role, directing the acyl-CoA product to a specific metabolic fate. **¶** Knowing the full complement of ACS genes in the human genome will facilitate future studies to characterize their specific biological functions.—Watkins, P. A., D. Manguel, Z. Jia, and J. Pevsner. Evidence for 26 distinct acyl-coenzyme A synthetase genes in the human genome. *J. Lipid Res.* 2007. 48: 2736–2750.

**Supplementary key words** fatty acid activation • fatty acid metabolism • conserved motifs • bioinformatics • consensus sequence • phylogenetic analysis • structure-function

Fatty acids serve many essential functions in living organisms. They are the building blocks of many lipids, including storage molecules such as triacylglycerol and cholesteryl esters, structural lipids such as phospholipids, plasmalogens, and sphingolipids, and signaling molecules such as diacylglycerol. Fatty acids can also be degraded for energy production, converted to alcohols or aldehydes, remodeled by elongation and/or insertion or removal of double bonds, or covalently bound to proteins. All of these metabolic processes have a common initial step, the “activation” of

the fatty acid by forming a thioester with CoA. This reaction is catalyzed by the acyl-coenzyme A synthetases (ACSs; EC 6.2.1.x) (1).

The diversity of fatty acids in nature is extensive. Fatty acids can range widely in their chain lengths, from the 2 carbon acid, acetate, to those containing >30 carbons in some waxes and plant lipids. Furthermore, fatty acids can be found that are totally saturated, that contain one (mono-unsaturated) or more (polyunsaturated) double bonds, or that have methyl branches. Thus, hundreds of naturally occurring fatty acid species exist. It is not surprising, therefore, that higher organisms contain multiple enzymes with ACS activity to facilitate both anabolic and catabolic reactions of fatty acids.

Before the era of abundant bioinformatic data, fatty acid activation activity was often characterized biochemically by chain length preference. The ACS activities found in different tissues and in different subcellular locations, particularly mitochondria and endoplasmic reticulum membranes (microsomes), were also characterized. These early investigations gave rise to the notion that there was an acetyl-CoA synthetase, a butyryl-CoA synthetase, a medium-chain ACS (ACSM), and a long-chain ACS (ACSL) (2–5). Subsequent studies predicted the existence of a very long-chain ACS (ACSVL) (6). However, there is significant overlap in the chain length specificity of, for example, ACSMs and ACSLs, and this may also vary depending on the degree of unsaturation of the fatty acid substrate. Nonetheless, classification provides a useful framework for defining subfamilies of related enzymes. Short-chain ACSs (ACSSs) typically activate acetate, propionate, or butyrate. ACSMs are those that activate C6 to C10 fatty acids. ACSLs are

Abbreviations: ACS, acyl-coenzyme A synthetase; ACSAc, yeast or bacterial acetyl-coenzyme A synthetase; ACSBG, bubblegum ACS; ACSF, ACS family; ACSL, long-chain ACS; ACSM, medium-chain ACS; ACSS, short-chain ACS; ACSVL, very long-chain ACS; BLAST, Basic Local Alignment Search Tool; EST, expressed sequence tag; FATP, fatty acid transport protein; HUGO, Human Genome Organization; LCFA, long-chain fatty acid; NCBI, National Center for Biotechnology Information; tACS, *Thermus thermophilus* ACS; VLCFA, very long-chain fatty acid.

<sup>†</sup>To whom correspondence should be addressed.

e-mail: [watkins@kennedykrieger.org](mailto:watkins@kennedykrieger.org)

<sup>S</sup>The online version of this article (available at <http://www.jlr.org>) contains supplementary data in the form of 3 Tables.

Manuscript received 9 February 2007 and in revised form 23 August 2007.

Published, JLR Papers in Press, August 30, 2007.

DOI 10.1194/jlr.M700378-JLR200

typically thought of as those that activate palmitate (C16:0) and oleate (C18:1), the most common fatty acids found in nature. However, the optimal chain lengths for these enzymes are frequently shorter, such as C12:0 (5, 7). ACSVLs have been so named not necessarily because they prefer very long-chain fatty acid (VLCFA) substrates but because they are capable of using these substrates. These enzymes often have a higher rate of activation of long-chain fatty acids (LCFAs) than VLCFAs (8–10).

With the sequencing of the human (and other) genomes completed, it is now possible to identify the entire complement of an organism's ACS genes and predicted protein products. Using highly conserved amino acid sequence motifs, 26 proven or likely human ACS genes were detected. Some have been characterized biochemically, whereas others have not yet been investigated. Two new candidate human ACSs were found to have enzymatic activity. Examination of amino acid sequences of all identified ACSs revealed conserved residues predicted by structural or biochemical studies to be important for catalysis and/or substrate binding. The availability of this information will facilitate future studies to elucidate the specific metabolic function of each ACS.

## MATERIALS AND METHODS

### Identification of candidate human ACS genes and proteins

Two highly conserved amino acid sequences, termed Motif 1 (20 amino acids) and Motif 2 (44–45 amino acids), were described previously for a group of 57 ACS sequences from diverse organisms, including mammals, fruitflies, roundworms, yeast, and bacteria (11). The Basic Local Alignment Search Tool (BLAST) algorithm (<http://www.ncbi.nlm.nih.gov/BLAST>) was used to probe the human nonredundant protein (using BLASTP and position-specific iterated BLAST) and human nucleotide (using TBLASTN) databases. This search strategy included queries consisting of *a*) four permutations of the central 10 amino acids of Motif 1 (YTSGTTGLPK, FTSGTTGLPK, YTSSTGLPK, and FTSGSTGLPK) and *b*) entire Motif 2 sequences of single representative human ACSS, ACSM, ACSL, and ACSVL and the human "bubblegum" (ACSBG1) protein (11). To facilitate the identification of potential paralogs, the stringency of the statistical significance threshold (Expect parameter) was decreased by changing the default value of 10 to 1,000–10,000. Full-length amino acid sequences of previously unidentified candidate ACS sequences identified in the initial screen were used as queries in subsequent BLAST searches. All candidate ACS genes reported herein were identified by this method.

Two pattern searches were also performed with the following query sequence derived from Motif 2: T-G-D-x(6,8)-G-x(3)-[F, I, V, M]x(2)-R-x(4)-[I, I, F, V]x(3,4)-G-x(2)-[I, I, V, F]x(4)-[V, I, I]-E. The Protein Information Resource Pattern/Peptide Match program (<http://pir.georgetown.edu/pirwww/search/pattern.shtml>) was used to interrogate the Protein Information Resource nonredundant reference protein database of human proteins, and pattern-hit-initiated BLAST was used to probe the human nonredundant protein database. These searches yielded no additional candidate ACSs. In the syntax used for the above query sequence and for consensus sequences, any amino acid shown in square brackets can occupy that position. X(n) indicates a sequence of n unspecified amino acid residues.

### Validation of candidate human ACS genes and proteins

To verify that previously unidentified candidate sequences *a*) represented likely ACSs and *b*) represented proteins that were expressed in humans, several analyses were performed. First, the overall size of the predicted protein was considered, as most ACSs consist of 600–700 amino acid residues. Second, amino acid sequences were examined for the presence of the aforementioned conserved residues (Motifs 1 and 2), and the positions of these motifs within the sequences were determined (see Results). Third, putative open reading frames were used in BLAST searches of the mammalian nonredundant protein database to identify potential orthologs in nonhuman species. Fourth, BLAT (BLAST-Like Alignment Tool; <http://genome.ucsc.edu>) was used to search the human genome to ascertain whether the sequence was supported by genomic evidence. Fifth, a human expressed sequence tag (EST) database at the National Center for Biotechnology Information (NCBI) was queried to determine whether sequences supporting the expression of putative ACS genes were present. With the exception of one protein tentatively identified as amino adipic semialdehyde dehydrogenase [ACS family 4 (ACSF4)], candidate sequences not meeting these criteria were not considered further.

### Identification of additional conserved motifs and derivation of consensus sequences

Data from the crystal structures of two bacterial ACSs (12, 13) and one yeast ACS (14) were used to identify regions other than Motifs 1 and 2 that potentially contained conserved amino acid sequences. For all protein sequences, regions of interest containing up to 30 amino acids were chosen based on their locations relative to Motifs 1 and 2 and aligned using ClustalW (<http://www.ebi.ac.uk/clustalw>), MAFFT (15) (<http://www.ebi.ac.uk/mafft/>), or MUSCLE (16) (<http://www.ebi.ac.uk/muscle/>). Aligned sequences were further analyzed by generation of sequence logos (17) using WebLogo (18) (<http://weblogo.berkeley.edu>). Consensus sequences were derived from multiple sequence alignments and sequence logos.

### cDNA cloning and protein expression

A plasmid containing full-length human ACSF2 was purchased from OriGene (TrueClone collection, catalog number TC108350). For transfer to the mammalian expression vector, pcDNA3 (Invitrogen), the open reading frame was amplified by PCR in a two-step process. The first amplification used TC108350 as template and forward oligonucleotide 5'-aattggatccagagccatggctgtctactgc-3', which incorporates a *Bam*HI restriction site (underlined), and reverse oligonucleotide 5'-gttcccggaccatccaggag-3'. The PCR product was used as template for a second round of amplification with the same forward oligonucleotide and reverse oligonucleotide 5'-aaattgcgccgctattcacagattagatg-3', which incorporates a *Not*I restriction site (underlined). After digestion with restriction enzymes, the resulting 1,847 bp fragment was cloned into the *Bam*HI and *Not*I sites of pcDNA3. The nucleotide sequence was determined and found to be identical to that of NM\_025149. Human ACSF3 full-length cDNA was also amplified by PCR using a human liver cDNA library (Clontech) as template and forward oligonucleotide 5'-cccgaattccttacctctctctctggct-3', which incorporates an *Eco*RI site (underlined), and reverse oligonucleotide 5'-ggatctagacgtgttctcgggtgtgaagg-3', which incorporates an *Xba*I site (underlined). After restriction enzyme digestion of the PCR product, the 1,865 bp fragment was cloned into the *Eco*RI and *Xba*I sites of pcDNA3 and completely sequenced. The sequence was identical to that of NM\_174917.

## ACS assays

COS-1 cells (American Type Culture Collection) were transfected with full-length cDNA constructs encoding ACSF2, ACSF3, or the empty pcDNA3 vector by electroporation as described previously (10). Three days after transfection, cells were harvested and subjected to at least one freeze-thaw cycle as described (10) and assayed for their ability to activate [ $^{14}\text{C}$ ]octanoic acid (American Radiolabeled Chemicals), [ $^{14}\text{C}$ ]palmitic acid (Moravek Biochemicals), or [ $^{14}\text{C}$ ]lignoceric acid (Moravek Biochemicals) as described previously (19). Final fatty acid concentrations in assays were 400  $\mu\text{M}$  for octanoate and 20  $\mu\text{M}$  for palmitate and lignocerate and included  $\sim 100,000$  dpm (1 nmol) of labeled fatty acid. Fatty acids were solubilized using  $\alpha$ -cyclodextrin (10 mg/ml in 10 mM Tris, pH 8.0) and incubated for 20 min at 37°C in 40 mM Tris, pH 7.5, 10 mM ATP, 10 mM  $\text{MgCl}_2$ , 0.2 mM CoA, 0.2 mM DTT, and cell suspension (15  $\mu\text{g}$  protein/assay for octanoate and palmitate, 60  $\mu\text{g}$  protein/assay for lignocerate) in a total volume of 250  $\mu\text{l}$ . For assay of ACSF3, reaction mixtures also contained Triton X-100 (final concentration, 0.1%). Assays were terminated by the addition of ice-cold Dole's solution, and separation of aqueous (acyl-CoA) and organic (fatty acid) phases was done according to the method of Dole (20). Radioactivity in the aqueous phase was quantitated by scintillation counting.

## ACS nomenclature

Many inconsistencies in the names of the various ACS enzymes can be found in the literature, some of which have affected the approved names for genes encoding these proteins. Recently, a consensus was reached between investigators and the Human Genome Organization (HUGO) Gene Nomenclature Committee regarding the mammalian ACSLs (21). The gene name "acyl-CoA synthetase long-chain family member\_" was approved. Approved gene symbols "ACSL\_" have a hierarchical structure, with the root "ACS" followed by the letter "L" for long-chain and a number designating each subfamily member. In part based on findings reported here, similar changes were approved in 2005 for some, but not all, members of the ACSS, ACSM, and ACSBG subfamilies.

The rationale for the ACSS family nomenclature is as follows. The older approved gene symbols for the two known human acetyl-CoA synthetases were ACAS2 and ACAS2L (for ACAS2-like); there was no ACAS1. Converting to the uniform nomenclature system for the ACS proteins, ACAS2 became ACSS2 and ACAS2L became ACSS1. The proposed name for a third subfamily member, identified in this work, is ACSS3. The ACSM nomenclature changes were based on the following rationale. The protein encoded by the BUCS1 gene was also referred to as MACS1 (22), which was changed to ACSM1 in the uniform nomenclature system. Another well-described protein was the HXM-A form of xenobiotic/medium-chain fatty acid:CoA ligase (23). Its gene name was changed by HUGO first to ACSM2 and, more recently, to ACSM2B. The human SA protein has been designated ACSM3. Three additional ACSM subfamily members are described in the present study. The proposed name for the human protein most similar to an olfactory-specific ACSM described in rats is ACSM4. The proposed names for the two remaining human ACSM candidates are ACSM5 and ACSM6. However, because of the high sequence similarity between ACSM2B and ACSM6, the latter was recently renamed ACSM2A by HUGO. The bubblegum ACS first reported in *Drosophila* (24) and later in humans (11) was designated ACSBG1, and a second homolog was designated ACSBG2 (25).

The approved gene names and symbols for the six members of the ACSVL subfamily have not yet been changed. These proteins were also reported to be fatty acid transport proteins

(FATPs) (26, 27); thus, their approved gene names and symbols are "solute carrier family 27 (fatty acid transporter) member\_" and "SLC27A\_" respectively. We propose that the first enzyme described as being capable of activating VLCFAs (28, 29), currently SLC27A2, be designated ACSVL1. We suggest that SLC27A6, the protein with the highest amino acid identity to ACSVL1, be called ACSVL2. We propose that SLC27A3 be named ACSVL3, SLC27A1 (FATP1) be named ACSVL4, and SLC27A4 (FATP4) be named ACSVL5. Finally, we suggest that SLC27A5, which preferentially activates the acyl side chain of bile acids rather than fatty acids (30, 31), be called ACSVL6.

Four proteins identified herein could not be assigned to the ACSS, ACSM, ACSL, ACSVL, or ACSBG subfamily. All have structural features suggesting that they belong to the greater ACS family. HUGO nomenclature advisors have suggested using the interim designation ACSF (for ACS Family) members 1–4.

## Phylogenetic analysis

ACS sequences for five additional species (the mouse *Mus musculus*, the zebrafish *Danio rerio*, the fruitfly *Drosophila melanogaster*, the nematode *Caenorhabditis elegans*, and the yeast *Saccharomyces cerevisiae*) were identified using a limited subset of the criteria used to identify human ACS sequences. BLAST searches used YTS GTTGLPK and FTS GTTGLPK as query sequences, and only matches with provisional or final RefSeq entries in NCBI databases were considered further. Amino acid sequence alignment was performed using MUSCLE and MAFFT; if Motif 2 was absent, the sequence was discarded. Although this method was not exhaustive, 113 likely ACS sequences from non-human species were identified. Phylogenetic trees were generated using the PAUP (Phylogenetic Analysis Using Parsimony) (32) and MEGA4 (Molecular Evolutionary Genetic Analysis) (33) programs. For neighbor-joining analysis in MEGA4 (34), evolutionary distances were computed using the Poisson correction method. All positions containing gaps were eliminated from the data set (complete deletion option). For the 139 proteins (26 human plus 113 nonhuman) shown in Fig. 3 below, there were 175 total positions in the final data set. The robustness of tree topology was evaluated by bootstrap analysis using a resampling size of 1,000 replicates. Segregation of ACSs into families was done by phylogenetic analyses of supported clades (see Results).

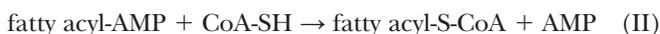
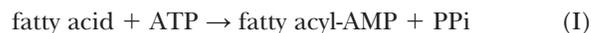
## Other methods

Many properties of human ACSs were obtained from publicly available databases, such as NCBI (<http://www.ncbi.nlm.nih.gov>), University of California Santa Cruz Genome Bioinformatics (<http://genome.ucsc.edu>), and Ensembl ([http://www.ensembl.org/Homo\\_sapiens](http://www.ensembl.org/Homo_sapiens)). BLAT was used to determine the number of exons encoding each ACS gene using the NCBI nucleotide reference sequences. For ACS activity measurements, statistical significance was determined using Student's *t*-test.

## RESULTS

### Refining the conserved sequence motifs used to identify candidate human ACSs

The process of fatty acid activation catalyzed by ACSs consists of two half-reactions (35, 36):



The first step is an ATP-dependent adenylation of the substrate with concomitant release of pyrophosphate (reaction I). CoA-SH then displaces the adenylyl group, which is released as AMP, and forms a thioester bond with the substrate (reaction II). Previous work from us and others indicated that proteins with ACS activity contain at least two distinctive and highly conserved amino acid sequence motifs, which likely serve important functions in substrate binding and/or catalysis. The first of these is a sequence of 10 amino acids, beginning approximately at residue 260 (see below), that is extremely well conserved in ACSs from bacteria to humans (11, 37, 38). This sequence, herein designated Motif I (Table 1, Fig. 1A),<sup>2</sup> is found within the 11 amino acid putative AMP binding domain signature in the PROSITE database (PDOC00427, PS00455) (<http://expasy.org/prosite>). Initial screening of protein and DNA databases for ACS candidates used Motif I sequences, as described in Materials and Methods.

Black and coworkers (39) examined a group of ACSs from bacteria, yeast, and mammals and identified a second region of highly conserved amino acid sequence. On the basis of site-directed mutagenesis experiments, these investigators concluded that this "signature motif" promoted acyl chain length specificity. We previously examined amino acid sequences of 57 ACSs (including 39 mammalian proteins) and identified a region of 44–45 residues, which we called Motif 2, that partially overlaps with the signature motif and that could be used *a*) to identify candidate ACSs and *b*) to organize ACSs into groups or subfamilies (11). Found within both Motif 2 and the signature motif Black et al. (39) is an arginine residue that is nearly invariant in ACS sequences (Table 1). Using this residue as an anchoring point, we reevaluated 30 upstream and 30 downstream residues in the aforementioned 57 ACS proteins using multiple sequence alignment. This extended region includes both the signature motif and our previously described Motif 2. Within this sequence of 61 amino acids, several other positions located at similar distances from the anchoring Arg contain conserved residues. This analysis yielded a revised consensus sequence of 36–37 amino acids, herein designated Motif II:

T-G-D-X(7)-G-X(3)-[*h*]X(2)-R-X(4)-[*h*]X(3,4)-G-X(2)-[*h*]X(4)-[*h*]E

where *h* includes any of the hydrophobic residues I, L, V, F, and M, and X is any amino acid (Table 1).

In all previously documented ACS sequences, this second conserved domain was invariably located downstream of Motif I, with the conserved Arg ~260 residues from Motif I (see below). Therefore, to identify all human ACSs, we used both Motif II (Table 1, Fig. 1B) and the previously described longer sequences (11) in secondary and tertiary

<sup>2</sup> In our previous publications, motifs were designated by Arabic numerals (Motifs 1 and 2) (11, 61, 62). Because this work has led to the refinement of consensus sequences, motifs are now designated by Roman numerals.

screens of human protein and DNA databases for candidate ACSs, as described in Materials and Methods.

### Identification of new candidate human ACS genes and proteins

During the last several decades, many human and other mammalian ACS genes and their predicted protein products have been reported. These include enzymes capable of activating short-, medium-, long-, and very long-chain FAs and related acyl-containing compounds such as bile acids, bile acid precursors, and acetoacetate. All of these known ACSs were found to contain both Motif I and Motif II, and the relative positions of these domains within the coding sequences are as expected, with the conserved Arg of Motif II ~260 residues downstream of Motif I. We hypothesized that the human genome might encode additional proteins with ACS activity.

Our primary screen, which probed NCBI databases for the four most commonly encountered variants of Motif I, identified ~100 human proteins or protein fragments containing sequences with significant homology to Motif I of bona fide ACSs. Many of these proteins were previously identified as ACSs, and many redundant sequences were present. However, several new candidate proteins were also detected. Secondary BLAST searches using the Motif II sequences from representative human ACSs, ACSMs, ACSs, and ACSVLs, as well as from ACSBG1, as query sequences did not identify additional candidate genes or proteins. However, subsequent BLAST searches using full-length amino acid sequences of new candidate proteins as the query identified additional proteins or protein fragments containing potential Motif I and/or Motif II sequences. A tertiary screening using either the Pattern/Peptide Match program or pattern-hit-initiated BLAST revealed no additional new sequences.

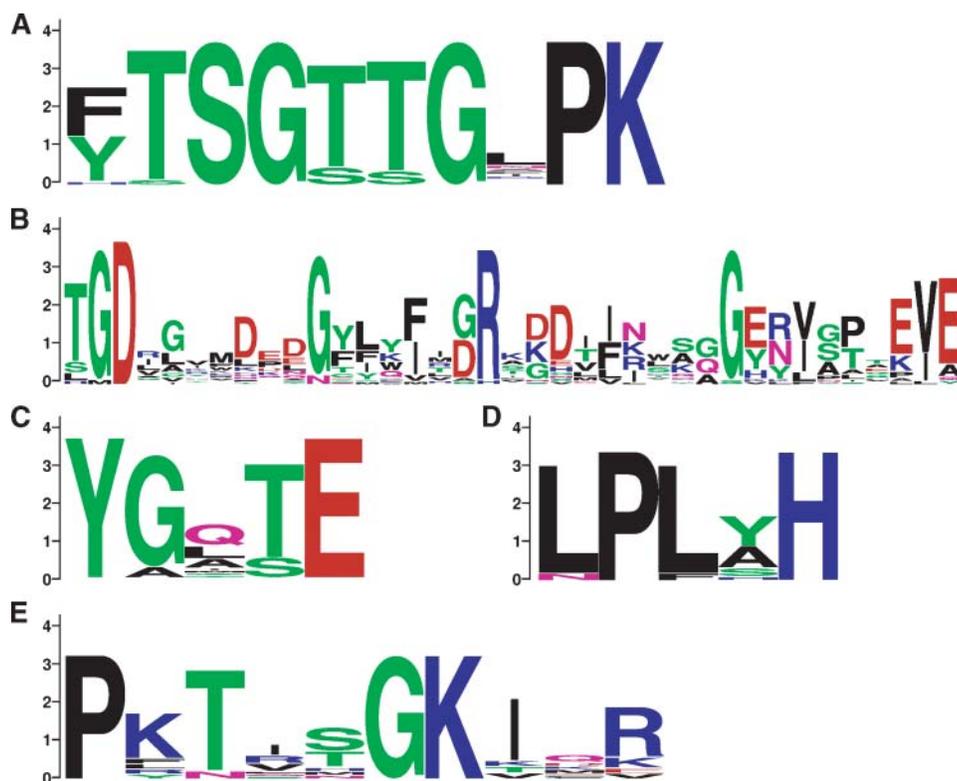
Using the above sequence analyses and related bioinformatics tools, we found a total of 26 ACS genes in the human genome (Tables 2, 3). Twenty of these genes encode proteins previously reported to have ACS activity in either humans or rodents, and six remain candidate ACS genes. The latter include genes provisionally designated *ACSM2A*, *ACSS3*, *ACSM5*, *ACSF2*, *ACSF3*, and *ACSF4* (Tables 2, 3). [Concurrent with this work, we established that murine *ACSF2* was enzymatically active (D. Maiguel, M. Morita, Z. Pei, M. L. Maguire, Z. Jia, and P. A. Watkins, unpublished observation).] The amino acid sequences of two of these candidate ACSs, *ACSM2A* and *ACSM5*, have been published because of their homology to *ACSM3* (formerly known as SAH) (40, 41). However, the biological functions of *ACSM2A* and *ACSM5* have not yet been characterized in any species.

The amino acid sequences of three candidate human ACSs (*ACSS3*, *ACSF2*, and *ACSF3*) have not been published previously. The sequence of rodent *ACSM4* has been published (42), but not that of the human protein. A multiple sequence alignment of these four proteins is shown in Fig. 2. The sequence of a known human ACSL, *ACSL4*, is included for comparison. Another of the candidate ACS genes, *ACSF4*, was reported to encode 2-

TABLE 1. Conserved amino acid sequence motifs in human ACSs

ACS	Motif I	Motif II	Motif III	Motif IV	Motif V
<b>Short-chain ACS family</b>					
ACSS1	293-YTSGTGMPPK-302	531-TGDGAYRTEGGYQQTGRMDDVINISGHRLGTAIE-566	443-WWQTE-447	Not found	636-PKTRSCGVKVMR-645
ACSS2	315-YTSGSTGKPK-324	550-TDGGCQRDQGGYWTGRIDDMLNVSGLLSTAEVE-585	465-FWQTE-469	Not found	655-PKTRSCGVKIMR-664
ACSS3	298-YTSGTTGLPK-307	537-TMDAGYMDDEEGLYVMSRVDDVINVAGHRISAGAIE-572	448-WWQTE-452	Not found	643-PKTRSCGVKIPR-652
<b>Medium-chain ACS family</b>					
ACSM1	225-FTSGTTGFPK-234	450-TGDRGKMDDEEGYGLGRSDDIINASGYRIGPAEVE-485	367-YGQSE-372	Not found	557-PKTTTGKIER-566
ACSM2A	220-FTSGTGLPK-229	444-LGDRCHKDEEGYQFMGRANDIINSSGYRIGPSEVE-479	361-YGQTE-365	Not found	551-PKTVTGTGKIQR-560
ACSM2B	220-FTSGTGLPK-229	444-LGDRCHKDEEGYQFMGRADDIINSSGYRIGPSEVE-479	361-YGQTE-365	Not found	551-PKTVTGTGKIQR-560
ACSM3	234-FTSGTSGPK-243	459-TGDRGYMDKDGFWFVARADDVILSSGYRIGPPEVE-494	376-YGQTE-380	Not found	565-PKTTTSGKTKR-574
ACSM4	228-FTSGTTGFPK-237	453-TGDRGYMDSGFWFVYGRADDVISSGYRIGPPEVE-488	370-TGQTE-374	Not found	560-PKTTTSGKTKR-569
ACSM5	229-FTSGTTGAPK-238	453-TGDRARMDKDGFWFVYGRADDVISSGYRIGPPEVE-488	370-YGQSE-374	Not found	560-PKTVTSGKIQR-569
<b>Long-chain ACS family</b>					
ACSL1	275-FTSGTTGNPK-284	538-TGDIGKWLPLNGTLKIDRRKKHFKLAQGEYIAPEKIE-574	458-YGQTE-462	320-LPLAH-324	675-(KAKR)-678
ACSL3	286-YTSGTGLPK-295	562-TGDIGEFEPDCLKIIDRKKDLVQLQAGEYVSLGKVE-598	475-TGLTE-479	328-LPLAH-332	699-(KLRK)-702
ACSL4	236-YTSGTGRPK-245	512-TGDIGEFHPDGLQIDRRKKDLVQLQAGEYVSLGKVE-548	425-YGLTE-429	278-LPLAH-282	649-(KLRK)-652
ACSL5	316-FTSGTTGDPK-325	579-TGDIGRWLPLNGTLKIDRRKKHFKLAQGEYIAPEKIE-615	499-YGQTE-503	361-LPLAH-365	716-(KAKR)-719
ACSL6	300-FTSGTTGNPK-309	563-TGDIGKWLPLNGTLKIDRRKKHFKLAQGEYVAPKIE-599	483-YGQTE-487	345-LPLAH-349	700-(KAKR)-703
<b>Very long-chain ACS family</b>					
ACSVL1	224-YTSGTTGLPK-233	463-SGDLMLMVDHENFIYFHDRLVGDTRWKGENVATTEVA-498	362-YAATE-366	264-LPFYH-268	572-(KHRK)-575
ACSVL2	223-FTSGTTGLPK-232	462-TGDLIVQDQDNFLYFWDRTGDFRWKGENVATTEVA-497	361-YAATE-365	263-LPLYH-267	571-(KLLK)-574
ACSVL3	334-FTSGTTGLPK-343	573-TGDLVQDDQGFLLRFHDDRTGDFRWKGENVATTEVA-608	472-YGLTE-476	374-LPLYH-378	682-(KQKQ)-685
ACSVL4	248-YTSGTTGLPK-257	490-SGDVLMDELGYMVMFRDRTGDFRWKGENVSTTEVE-525	387-YGATE-391	289-LPLYH-293	598-(KIQK)-601
ACSVL5	245-YTSGTTGLPK-254	487-TGDVLMDELGYLYFRDRTGDFRWKGENVSTTEVE-522	384-YGATE-388	286-LPLYH-290	595-(KFKQ)-598
ACSVL6	294-YTSGTTGLPK-303	533-TGDVLMAMDREGFLYFRDLGDFRWKGENVSTTEVE-568	432-YGSTE-436	334-LPLYH-338	642-(KLMK)-645
<b>Bubblegum ACS family</b>					
ACSBG1	281-YTSGTTGNPK-290	548-TGDAGRLDADGFLYITGRLLKELITAGGENVPPVPIE-584	474-YGLSE-478	326-LPLSH-330	701-(KLRK)-704
ACSBG2	229-YTSGTTGIPK-238	494-SGDLGQLDGLGFLYVTG#IKELLITAGGENVPPVPIE-530	420-YGLSE-424	271-LPLSH-275	647-(KIQK)-650
<b>Other ACSs</b>					
ACSF1	289-FSSGTTGAPK-298	521-HGDYCRINPKTGIVMLGRSDGTLNPNVRFSGSEIV-557	Not found	Not found	627-PYTLNGKKVE-636
ACSF2	262-FTSGTTGSPK-271	491-TGDVATMNEQGFKIVGRSKDMHIRGGENIYPAELE-526	408-YGTTTE-412	307-NPLYH-311	593-PLTISGKIQK-602
ACSF3	201-YTSGTTGPRPK-210	455-TGDTIVFKDGGYWRGRTSVDIIKTGGYKSALEVE-490	355-YGMTE-359	242-LPLHH-246	557-PRNQMGKIDK-566
ACSF4	197-HTSGTTGIPK-206	426-TGDFVTVKDGEIFFLGRKDSQJKRHRGKRLNIELVQ-460	344-YGITE-348	Not found	521-PFTSHGKIDV-530
Consensus	[Y,F]TSG[T,S]TGXPK	TGDX(7)GX(3)hX(2)RX(4)hX(3,4)GX(2)hX(4)hE	YGXTE(WWQTE)	LPLXH	PKTX[S,T]GKIX[R,K] (KXX[R,K])

ACS, acyl-coenzyme A synthetase; ACSBG, bubblegum ACS; ACSF, ACS family; ACSL, long-chain ACS; ACSM, medium-chain ACS; ACSV, short-chain ACS; ACSV, very long-chain ACS. Conserved domains were identified as described in the text. The position of each motif within the open reading frame is noted; for ACSs that have recognized transcript variants (Table 3), the motif position for variant 1 is shown. The presence of both Motifs I and II was the criterion used for the classification of human proteins as ACSs. Motif II contains a highly conserved arginine residue (R) shown in boldface. Motifs III and IV were not found in all candidate sequences. A variation of the Motif III consensus found in the ACS family is shown in parentheses. Motif V, a potential acetylation domain, is found in about half of the human ACS sequences. A fragment of Motif V (consensus shown in parentheses) is found in the remaining sequences. The lysine residue (K) that is potentially acetylated is shown underlined in boldface. Consensus sequences were derived from sequence logos (Fig. 1) and multiple sequence alignments. The symbol *h* indicates any of the hydrophobic amino acids (I, L, V, F, or M).



**Fig. 1.** Consensus sequences of human acyl-coenzyme A synthetase (ACS) motifs. The WebLogo program was used to generate sequence logos from which consensus sequences were obtained for Motif I (A), Motif II (B), Motif III (C), Motif IV (D), and Motif V (E), as described in Materials and Methods. These logos are graphic representations of amino acid sequences obtained from multiple sequence alignments. Although the overall height of each stack reflects the degree of conservation at that position, the height of each letter represents the relative frequency at which a given amino acid is found at that position (18). For Motif II, gaps were eliminated to facilitate visualization (see Table 1). Consensus for Motifs III, IV, and V were obtained using 22, 15, and 13 sequences, respectively, as described in the text. These consensus sequences are presented in Table 1.

aminoadipic 6-semialdehyde dehydrogenase (43). However, the predicted protein contains both Motif I and Motif II consensus sequences (Table 1). Therefore, alignment of the ACSF4 amino acid sequence is also included in Fig. 2.

With the exceptions of a transcript variant of ACSM3 (ACSM3\_v2) and ACSF4 (see below), all human ACSs were of the expected size, with an average length of 656 amino acids (range, 576–739) (Table 3). The relative positions of Motifs I and II were also as predicted from our previous study of ACSs from various species (11). The initial residue of Motif I was found, on average, at residue 261 (range, 201–334), and the average distance between this residue and the conserved Arg of Motif II was 263 amino acids (range, 241–293) (Table 1).

#### ACSM2A and ACSM2B are distinct genes

The *ACSM2A* and *ACSM2B* genes and their encoded proteins are nearly identical and thus difficult to distinguish. The coding sequences of these genes are 98.8% identical, and their amino acid sequences are 97.6% identical. Thus, it would be possible to infer that experimentally determined differences were attributable to polymorphisms or sequencing errors. However, there is ample

evidence supporting the existence of both genes. Both are located on chromosome 16p12.3, but whereas *ACSM2A* is on the plus strand, *ACSM2B* is on the minus strand (Table 3). Both nucleotide sequences are supported by genomic sequence data and the existence of informative ESTs (Table 3). *ACSM2A* and *ACSM2B* have 20 nucleotide differences in the coding region (involving 19 codons), of which 14 are nonsynonymous substitutions and 6 are synonymous substitutions. Of the amino acid changes resulting from nonsynonymous substitution, only one lies within a conserved motif. Residue 463, found in Motif II, is Asn in *ACSM2A* and Asp in *ACSM2B* (Table 1). Although the 3' untranslated regions of *ACSM2A* and *ACSM2B* are also very related (94.7% identity over 113 bp), the 5' untranslated regions of the two transcripts show more variability (59.6% identity over 146 bp). Despite these differences, distinguishing these genes/proteins experimentally (e.g., by Northern blot or Western blot) would be extremely difficult.

#### ACS transcript variants

Multiple isoforms of seven human ACSs were detected by BLAST searching. Two variants each of human *ACSM2*,

TABLE 2. Human acyl-CoA synthetase family nomenclature

Proposed Designation	Approved Gene Symbol	Approved Gene Name	Alternative Symbols and Aliases
<b>Short-chain ACS family</b>			
ACSS1	<i>ACSS1</i>	Acyl-CoA synthetase short-chain family member 1	ACAS2L; AceCS2L; MGC33843
ACSS2	<i>ACSS2</i>	Acyl-CoA synthetase short-chain family member 2	ACAS2; ACS; ACSA; AceCS; dJ1161H23.1
ACSS3			FLJ21963
<b>Medium-chain ACS family</b>			
ACSM1	<i>ACSM1</i>	Acyl-CoA synthetase medium-chain family member 1	BUCS1; MACS1
ACSM2A	<i>ACSM2A</i>	Acyl-CoA synthetase medium-chain family member 2A	LOC123876; A-923A4.1
ACSM2B	<i>ACSM2B</i>	Acyl-CoA synthetase medium-chain family member 2B	HXMA; HYST1046
ACSM3	<i>ACSM3</i>	Acyl-CoA synthetase medium-chain family member 3	SAH; SA
ACSM4			LOC341392
ACSM5			FLJ20581
<b>Long-chain ACS family</b>			
ACSL1	<i>ACSL1</i>	Acyl-CoA synthetase long-chain family member 1	ACS1; LACS; FACL1; FACL2; LACS1; LACS2
ACSL3	<i>ACSL3</i>	Acyl-CoA synthetase long-chain family member 3	ACS3; FACL3; PRO2194
ACSL4	<i>ACSL4</i>	Acyl-CoA synthetase long-chain family member 4	ACS4; FACL4; LACS4; MRX63
ACSL5	<i>ACSL5</i>	Acyl-CoA synthetase long-chain family member 5	ACS2; ACS5; FACL5
ACSL6	<i>ACSL6</i>	Acyl-CoA synthetase long-chain family member 6	ACS2; FACL6; LACS2; LACS5; KIAA0837
<b>Very long-chain ACS family</b>			
ACSVL1	<i>SLC27A2</i>	Solute carrier family 27 (fatty acid transporter), member 2	VLCS; FATP2; VLACS; FACVL1; hFACVL1; HsT17226
ACSVL2	<i>SLC27A6</i>	Solute carrier family 27 (fatty acid transporter), member 6	FATP6; FACVL2; VLCS-H1
ACSVL3	<i>SLC27A3</i>	Solute carrier family 27 (fatty acid transporter), member 3	FATP3; VLCS-3; MGC4365
ACSVL4	<i>SLC27A1</i>	Solute carrier family 27 (fatty acid transporter), member 1	FATP; FATP1; FLJ00336; MGC71751
ACSVL5	<i>SLC27A4</i>	Solute carrier family 27 (fatty acid transporter), member 4	FATP4
ACSVL6	<i>SLC27A5</i>	Solute carrier family 27 (fatty acid transporter), member 5	FATP5; FACVL3; VLACSR; VLCSH2; VLCS-H2; FLJ22987
<b>Bubblegum ACS family</b>			
ACSBG1	<i>ACSBG1</i>	Acyl-CoA synthetase bubblegum family member 1	BG1; BGM; FLJ30320; MGC14352
ACSBG2	<i>ACSBG2</i>	Acyl-CoA synthetase bubblegum family member 2	BGR; PRTD-NY3; DKFZp434K1635
<b>Other ACSs (ACS Family)</b>			
ACSF1	<i>AACS</i>	Acetoacetyl-CoA synthetase	SUR-5; FLJ12389
ACSF2			FLJ20920
ACSF3			LOC197322
ACSF4	<i>AASDH</i>	2-Aminoadipic 6-semialdehyde dehydrogenase	NRPS998; NRPS1098

*ACSM3*, *ACSL3*, *ACSL4*, *ACSL6*, and *ACSVL2* were found, along with three variants of *ACSL5* (Table 3). The *ACSL3* and *ACSVL2* variants, along with two of three *ACSL5* variants, differ only in their 5' untranslated regions and are expected to encode identical proteins. The *ACSS2* variants are predicted, according to NCBI RefSeq annotation, to arise via an alternative splicing event whereby a different first exon (which includes the initiator methionine codon) is used. The protein encoded by *ACSS2\_v2* is shorter than that encoded by *ACSS2\_v1* by 50 amino acids at the N terminus.

The *ACSL4* variants arise from the use of different exons 3 (each containing an initiator methionine). *ACSL4\_v2* also has an additional exon (exon 4) not found in *ACSL4\_v1*. The encoded proteins thus differ at their predicted N termini, with *ACSL4\_v2* containing 41 additional amino acids than *ACSL4\_v1*. The longer isoform is the predominant form in human brain (44). *ACSL5\_v1* is also longer than either *v2* or *v3* at the N terminus. Whereas exon 1 in *ACSL5\_v1* contains an in-frame ATG codon, exons 1 in the other two variants do not. The latter variants use an alternative in-frame start codon found in exon 2, according to NCBI RefSeq annotation.

*ACSL6* transcript variants are distinct from the other ACS variants identified in that they differ not at their N or C termini but internally. Two alternative exons 11 are found in the human *ACSL6* gene. Exon 11 contains the sequence encoding the conserved domain referred to as

Motif IV (see below). To date, the proteins encoded by rat *ACSL6\_v1* and *ACSL6\_v2* are the only variants whose enzymatic activity has been investigated in any species (45).

The NCBI database contains a reference sequence (NM\_202000) for a variant of *ACSM3* (*ACSM3\_v2*) whose predicted protein product would lack Motif II and thus would not satisfy our criteria for a candidate ACS. This variant is predicted to arise via an alternative splicing event that results in a longer exon 9, which contains a stop codon. We predict that *ACSM3\_v2* would not be an enzymatically active ACS, but we include it in Table 3 for the sake of completeness.

#### Two of the newly identified human ACS candidates are enzymatically active

As proof of principle, we chose two candidate ACSs identified in this screen for functional studies. We cloned full-length cDNA encoding human ACSF2 and ACSF3 and expressed the proteins in COS-1 cells. We then examined the ability of ACSF2- or ACSF3-overexpressing cells to activate a representative medium-, long-, or very long-chain FA. Compared with vector-transfected cells, ACSF2-expressing cells robustly activated the 8 carbon medium-chain fatty acid, octanoate, but not the 16 or 24 carbon substrate (Table 4). In contrast, ACSF3-expressing cells showed a preference for lignoceric acid, a 24 carbon VLCFA (Table 4). Neither ACSF2 nor ACSF3 showed significant ability to activate the 16 carbon LCFA, palmitate. Thus, both ACSF2 and ACSF3 are, as predicted, fatty acyl-CoA synthetases.

TABLE 3. Properties of human acyl-CoA synthetases

ACS	RefSeq Nucleotide	RefSeq Protein	UniGene	Gene ID	MIM	Chromosomal Location	Strand	Exons	Amino Acids	Supporting Sequences
ACSS1	NM_032501	NP_115890	Hs.529353	84532		20p11.23-p11.21	-	14	689	475
ACSS2_v1	NM_018677	NP_061147	Hs.517034	55902	605832	20q11.22	+	18	701	662
ACSS2_v2	NM_001076552	NP_001070020	Hs.517034	55902	605832	20q11.22	+	18	651	662
ACSS3	NM_024560	NP_078836	Hs.259559	79611		12q21.31	+	16	686	178
ACSM1	NM_052956	NP_443188	Hs.306812	116285		16p12.2	-	13	577	34
ACSM2A	NM_001010845	NP_001010845	Hs.298252	123876		16p12.3	+	15	577	218 (7)
ACSM2B	NM_182617	NP_872423	Hs.298252	348158		16p12.3	-	15	577	218 (9)
ACSM3_v1	NM_005622	NP_005613	Hs.647340	6296	145505	16p13.11	+	14	586	169
ACSM3_v2	NM_202000	NP_973729	Hs.160976	6296	145505	16p13.11	+	9	438	169
ACSM4	NM_001080454	NP_001073923	Hs.450804	341392		12p13.31	+	13	580	18
ACSM5	NM_017888	NP_060358	Hs.578946	54988		16p12.3	+	14	579	114
ACSL1	NM_001995	NP_001986	Hs.406678	2180	152426	4q34-q35	-	21	698	706
ACSL3_v1*	NM_004457	NP_004448	Hs.471461	2181	602371	2q34-q35	+	17	720	762
ACSL3_v2*	NM_203372	NP_976251	Hs.471461	2181	602371	2q34-q35	+	16	720	762
ACSL4_v1	NM_004458	NP_004449	Hs.268785	2182	300157	Xq22.3-q23	-	16	670	500
ACSL4_v2	NM_012297	NP_075266	Hs.268785	2182	300157	Xq22.3-q23	-	17	711	500
ACSL5_v1	NM_016234	NP_057318	Hs.11638	51703	605677	10q25.1-q25.2	+	21	739	461
ACSL5_v2*	NM_203379	NP_976313	Hs.11638	51703	605677	10q25.1-q25.2	+	21	683	461
ACSL5_v3*	NM_203380	NP_976314	Hs.11638	51703	605677	10q25.1-q25.2	+	21	683	461
ACSL6_v1	NM_015256	NP_056071	Hs.14945	23305	604443	5q31	-	21	722	182
ACSL6_v2	NM_001009185	NP_001009185	Hs.14945	23305	604443	5q31	-	21	722	182
ACSVL1	NM_003645	NP_003636	Hs.11729	11001	603247	15q21.2	+	10	620	255
ACSVL2_v1*	NM_014031	NP_054750	Hs.49765	28965	604196	5q23.3	+	11	619	59
ACSVL2_v2*	NM_001017372	NP_001017372	Hs.49765	28965	604196	5q23.3	+	10	619	59
ACSVL3	NM_024330	NP_077306	Hs.438723	11000	604193	1q21.3	+	10	730	306
ACSVL4	NM_198580	NP_940982	Hs.363138	376497	600691	19p13.11	+	12	646	145
ACSVL5	NM_005094	NP_005085	Hs.495227	10999	604194	9q34.11	+	13	643	203
ACSVL6	NM_012254	NP_036386	Hs.292177	10998	603314	19q13.43	-	10	690	203
ACSBG1	NM_015162	NP_055977	Hs.250616	23205		15q23-q24	-	14	724	170
ACSBG2	NM_030924	NP_112186	Hs.465720	81616		19p13.3	+	15	666	126
ACSF1	NM_023928	NP_076417	Hs.169054	65985		12q24.31	+	18	672	424
ACSF2	NM_025149	NP_079425	Hs.288959	80221		17q21.33	+	16	615	492
ACSF3	NM_174917	NP_777577	Hs.461727	197322		16q24.3	+	11	576	255
ACSF4	NM_181806	NP_861522	Hs.104347	132949		4q12	-	15	1098	136

All human ACSs identified in this report are listed, along with transcript variants for which Reference Sequences (RefSeq) are found in National Center for Biotechnology Information (NCBI) databases. Variants that differ only within their 5' untranslated regions are indicated by asterisks. UniGene, Gene ID, and MIM (Mendelian Inheritance in Man) identifiers were from NCBI databases. Chromosomal location and strand, exons, and number of amino acids were deduced from NCBI data and using the BLAST-Like Alignment Tool, as described in Materials and Methods. The number of database sequences supporting each gene (which includes all variants of a given gene) is shown. Although the NCBI attributes many supporting sequences to the highly similar ACSM2A and ACSM2B, a smaller number of sequences (shown in parentheses) provide unequivocal evidence that these are separate genes.

### Assignment of human ACS proteins into subfamilies

Using Expect values from BLAST and pairwise alignments as criteria, protein products of the 26 homologous candidate human ACS genes were assigned to subfamilies likely to represent ACSS, ACSM, ACSL, and ACSVL as well as ACSBG homologs (Table 1). Twenty-two proteins could be assigned to these subfamilies. The sequences of four proteins (ACSF1–ACSF4), particularly in their Motif II region, differed sufficiently from the other 22 proteins and from each other that they were not assigned to a specific subfamily. Phylogenetic analyses (see below) lent further support to these subfamily assignments.

Pairwise alignments of full-length amino acid sequences of these proteins revealed that amino acid identities between subfamily members ranged from 29% to 96% [averaging  $48 \pm 13\%$  (mean  $\pm$  SD) for 44 alignments], whereas the identity of nonsubfamily pairs ranged from 15% to 27% (averaging  $20 \pm 2\%$  for 281 alignments) (see supplementary Table I). Although the average percentage identity within subfamilies was higher for Motif II ( $71 \pm 13\%$  for 44 alignments) than for the full-length sequences,

there was a broader range (39–97%). The average identity of nonsubfamily pairs ( $32 \pm 7\%$  for 281 alignments) was slightly higher than that for full-length sequences, and the values ranged from 14% to 53% (see supplementary Table II). The highest degree of intersubfamily Motif II identity was observed between the ACSS and ACSM proteins, for which the range was 42–53%.

### Phylogenetic analyses of human and nonhuman ACS sequences

We performed phylogenetic analyses to infer the evolutionary relationships of the human proteins. Twenty-six human proteins were multiply aligned with MUSCLE; alignments with ClustalW or MAFFT differed somewhat (as is expected given their varying strategies for optimizing alignments) but yielded comparable phylogenetic trees (data not shown). We obtained comparable results using the neighbor-joining distance-based algorithm as well as maximum parsimony, and we also obtained similar results for the relationships of *a*) 26 human paralogs or *b*) these 26 human paralogs analyzed together with 113 orthologs identified in

ACSS3	MKPSWLQCRKVTSAAGLGGPLPGSSPARGAGAAALRALVVPGRGGLGGRGCRALSSGSGS	60
ACSM4	MKIFFRYQTFRFI-----WL-TKPP-----GRRLHKDQHL-	29
ACSF2	MAVYVGMRLRLGRLCAGSSGVLGARAALSRSWQEARLQ-----GVRFLSSREVD	48
ACSF3	-----MLPHVVLTFRRLL-----GCALAS-----	18
ACSF4	-----MTLQEL-----VHKAAS-----	12
ACSL4	-----MAKRIKAKPTSDKP-----GSPYRSVTHFD	25
ACSS3	EYKTHFAASVTDPERFWGKAAEQISWYKPWTKTLENKHSPTSRWFVEGMLNICYNAVDRH	120
ACSM4	-WTPLTLADFEAINR-----C-----NRPLPKNFNFAADVLDQWSQKEKTGER-----	71
ACSF2	RMVSTPIGGLSYVQG-----C-----TK---KHLNSKTVVQCLETTAQRVPER-----	88
ACSF3	-----C-----RLAPARHRGSGLLHTAPVARSDRSAPVFT-----	48
ACSF4	-----C-----YM-----DRVAVCFDECN--	26
ACSL4	SLAVIDIPGADTLDKLFDHAV-----SKFGKKDSLGTREILSEENEM-----	67
ACSS3	IENGKGDKIAIIYDSPVTNTKATFTTYKEVLEQVSKLAGVLVKHGI-KKGDVTV--IYMPM	177
ACSM4	-----PANPALWVWNGKGVKWSFRELGSLSRKAANVLTKPCGLQRGDRLA--VILPR	123
ACSF2	-----EALVVLH-----EDVRLTFAQLKEEVDAASGLLSIGL-CKGDR---LGMWG	131
ACSF3	-----RALAFGDRIALVDQHRHTYRELYSRSLRLSQEI-CRLCGCVGGDLREERSVFLC	102
ACSF4	-----NQLPVY-----YTYKTVVNAASELSNFLLLHCD-FQG-IRE--IGLYC	65
ACSL4	-----QPNGKVFKKLILGNKWMNYLEVNRVNNFGSGLTALGLPKNT-----IAIFC	116
ACSS3	IPQAMYTM--LACARIG----AIHSLIFGGFASKELSSRIDHVKPKVVVTSFGEIPEGR	230
ACSM4	IPE--WVLNVACIRTG----IIFMPGTIQLTAKDILYRLRASKACIVAS-----	168
ACSF2	PNSYAWVLMQLATAQAG----IILVSVNPAYQAMELEVVLKKGCKALVFP-----	178
ACSF3	ANDASYVVAQWASWMSG----GVAVPLYRKHFAAQLEYVICDSQSSVVLAS-----	149
ACSF4	QEG----IDLPSWILGILQVPAAYVPIEPDPSPLSTHFMKKCNLKYLLVEKKQINKPK	120
ACSL4	ETRAEWMIAAQTCFKYN----FPLVTLTYATLKEAVVHGLNESEASYLITS-----V	164
ACSS3	RVEYVPLVEEALKIGQHKPKDKILYINRPNMEAVPLAPGRD----LDWDEEMAKAQSHDC	285
ACSM4	--EEVAPAVESIVL-ECPLDKTKLLVSPQSWNGWLSFQEL-----FQFASSEHSC	215
ACSF2	KQFKTQQYYNVLKQ-ICEVENAQPGALKSQRLPDLTTVISVDAPLPGLLLEDDVVAAGS	237
ACSF3	-QEYLELLSPVVRK-----LGVPLLP-----LTPAIYTG	178
ACSF4	SPHETLLNYDTPTV-EHNDLVLFRLHWKNTVNLMLNDGK----EKYEKIKSISSEH	174
ACSL4	ELLESKLTALLDI-SC--VKHIIYVDNKAINKAEPYPEGF-----EIHMSQSVVEELGS	214
Motif I		
ACSS3	VPVLSEHP-----LYILYTSGTTGLPKGVIRPTGGYAVMLHWSMSS--IYGLQ	331
ACSM4	VETGSQEP-----MTIYFTSGTTGFPKMAQHSQSSNLGIGFTLCGRY--WLDLK	261
ACSF2	TRQHLDLQLQYNQQFLSCHDPINIQFTSGTTGSPKQATLSHYINVNNSNLGERLKLHEKT	297
ACSF3	VEEPAEVPVPEQGWNRNG--AMIYTSGTTGRPKGVLSTHQNIRAVVTGLVHK--W-AWT	233
ACSF4	VNEEKAEHMDLRKHKCL--AYVLHTSGTTGIPKIVRVPKCIIVPNIQH-FRV--LFDIT	229
ACSL4	NPENLGIPPSRPTPSDM--AIVMYTSGSTGRPKGVMMHHSNLIAGMTQCER--IPGLG	269
Motif IV		
ACSS3	PGEV--WAAASDLGWVGHHSYICYGPLLH-GNTTVLYEGKPVGTPDAGAYFRVLAEH--GV	387
ACSM4	SSDI-IWNMSDTGWVKAAIGSVFSSWLC--GACVFVHRMAQFDTDT--FLDTLTTY--PI	314
ACSF2	PEQLRMILPNPLYHCLG--SVAGTMMCLMYGATLILASPIFNGKK--ALEAISRE--RG	350
ACSF3	KDDV-ILHVLPLHHVHGUVNALLCPLWV-GATCVM--MPEFSPQQ-VWEEFLSSETPRI	287
ACSF4	QEDV-LFLASPLTFDPSVVEIFLA-LSS-GASLLIVPVSVKLLPSK-LASVLFSSH--RV	283
ACSL4	PKDT-YIGYLPPLAHVLE---LTAEISCTYTCRIGYSSPLTLDSDQ--SSKIKKGS--KG	320

**Fig. 2.** Alignment of newly identified human ACS amino acid sequences. Predicted sequences of human ACSS3, ACSM4, ACSF2, ACSF3, and ACSF4 were aligned along with a representative long-chain ACS, ACSL4, using MUSCLE, as described in Materials and Methods. The positions of Motifs I–V are indicated, and residues conforming to the consensus sequences shown in Table 1 are shaded. The sequence of ACSF4, which contains 1,098 amino acids, was truncated at 566 residues.

five additional species: mouse, zebrafish, fruitfly, nematode, and yeast. A neighbor-joining tree with all 139 homologs is shown in Fig. 3. The 26 human proteins are indicated by arrows and labels with larger font size. We observed 10 major groups, including five clades corresponding to ACSS, ACSM, ACSL, ACSVL, and ACSBG proteins. We also observed five clades that we designated ACSF1, ACSF2, ACSF3, ACSF4, and worm/fly. We performed 1,000 bootstrap replicates to assess the robustness of each node and observed strong support for the topology of the tree.

We noted several features of the phylogenetic tree shown in Fig. 3. First, the worm/fly clade lacked human, mouse, fish, or yeast members and thus appears to represent a nonvertebrate, nonfungal expansion. The worm/fly clade consisted of two subgroups (one set of 11 *Dro-*

*sophila* proteins forming a clade with 98% bootstrap support, and a second group of seven *Drosophila* and *C. elegans* proteins). Second, the medium-chain clade had six human members but no *Drosophila* or *C. elegans* members, and additional BLAST searching revealed no apparent medium-chain family members in other insect or nematode species. These species differences may reflect different metabolic requirements of these organisms. Five of the six human ACSM genes in the medium-chain clade are located on chromosome 16p12.2-13.11 and thus may have arisen by tandem duplication. Third, of the 10 clades we outlined in Fig. 3, most had an ancestral node that had  $\geq 95\%$  bootstrap support, indicating a robust estimate of the topology. The ACSF4 clade included an ancestral node with  $>95\%$  support, with a *Drosophila* protein (dm11) placed as an

ACSS3	AALFTAPTAI-----RAIRQQDPG-----AALGKQYSLT-----	416
ACSM4	TTLCSPTTVY-----RMLVQKDLK-----RYKFK-----	338
ACSF2	TFLYGTPTMFV-----DILNQPDFS-----SYDIS-----	375
ACSF3	NVFMVPTIYTKLMEYYDRHFTQPHAQ-----DFLRA-----	319
ACSF4	TVLQATPTLL-----RRFGSQLIK-----STVLS-----	307
ACSL4	DCTVLKPTLMAAVPEIMDRIYKYNVMSKVQEMNYIQKTLFKIGYDYKLEQIKKGYDAPLCN	380
Motif III		
ACSS3	-----RFKTLFVAGERCD-VETLE-WSKNVFRVPLDHHWQTEETGSPITASC	461
ACSM4	-----SLRHCLTGGEP LN-PEVLEQWRVQTGL-ELYEGYQTE-----VGMIC	379
ACSF2	-----TMCGGVIAGSPAP-PELIRAI INKINMKDLVVAYGTE-NSPVTFAH	420
ACSF3	-----VCEEKIRLMVSGSAALP-LPVLEKWKNI TGH-TLLERYGMTE-----IGMAL	364
ACSF4	-----ATTSRLRVLALGGEAFPSLTVLRSWRGEGNKQTQIFNVYGI TEVSSWATIYR	357
ACSL4	LLLFKKVKALLGGNVRMMLSGGAPLS-PQTHR-FMNVCFCCPIGQGYGLTE-----SC	431
ACSS3	VGL--GN--SKTPPPQAGKSPGVNVMILD-----DNMQKLRCL-	499
ACSM4	ANQKQGE--IK--PGSMGKMLPYDVQIID-----ENGNVLPPEKE-	416
ACSF2	FPEDTVE--QK--AESVGRIMPHTEARIMN-----MEAGTLAKLNTP-	458
ACSF3	SGPLTTA--VR--LPGSVGTPLPGVQVRIVSENPQREACSYTIHABGDERGRTKVPGFEB	420
ACSF4	IPEKTLNSTLKCCELPVQLGFPLGTVEVRD-----TNGFTIQEGS--	398
ACSL4	GAGTVTE--VTDYTTGRVGAPLICCEIKLKD-----WQEGGYTINDKPN	473
Motif II		
ACSS3	--GNIVVKLPL--PPGAFSGLWKNQEA FKHL YFEKFPG--YYDTMDAGYMDEEGYLYVMS	553
ACSM4	--GEIALRLKPTRPFCFFSKYVDNPQKTAATIRGD---FYVTGDRGMDSG YFFVVG	469
ACSF2	--GELCIR-----GYCVMLGYWGEPEKTEEAVDQDK---WYWTGDVATMNEQGFCKIVG	507
ACSF3	KEGELLVR-----GPSVFREYWNKPEETSAFTLDG---WPKTGDTVVF-KDQYWRIG	470
ACSF4	--GQVFLG-----GRNRVC--FLDDEVTVPLGTMR-----ATGDFVTV-KDGEIFFLG	441
ACSL4	PRGEIVIG----GQNI SMGYFKNEEKTAEDYSVDENGQRWFCTGDI GEFHPDGCLQIID	528
Motif I		
ACSS3	RVDDVIN-VAGHRISAGAT EESILSHGTVADCAVVGKEDPLK-----GHVPLALCVL	604
ACSM4	RADDV I I-SSGYRIGPFEVESALIEHPAVVESAVVSSPDQIR-----GEVVKAFVVL	520
ACSF2	RSKDMI I-RGGENIYPAELEDFFHHTHPKVQEVQVVGKDDRM-----GEEICACIRL	558
ACSF3	RTSVD I IKTGGYKVSAL E VEWHL LAHPSITDVAVIGVPMWTW-----GORVTA VVTL	522
ACSF4	RKDSQIK-RHGKRLNIELVQQVAEELQQVESC AV-----TW-----YNQEKLLLFMVS	488
ACSL4	RKKDLV K LQAG EYVSLGKVEAALKNCPLIDNI CAFAKSDQSVVISFVVVNPQKRLTL LAQQ	588
Motif V		
ACSS3	RK-----DINATEEQVLEEIVKHVRQ----NIGPVAAFRNAV FVKQLPKTRSG-----	648
ACSM4	AA-----PFKSYNPEKLTLELQDHVKK----STAPYKYPRKVEFVQELPKTITG-----	565
ACSF2	KD-----GEETTVEEIKAFCKG----KISHFKIPKYIVFVTVNYPLTISG-----	598
ACSF3	RE-----GHSLSHRELKEWARN----VLAPYAVPSELVVEEIPRNQMG-----	562
ACSF4	KD-----ASVKEYIFKELQK----YLP SHA VPDELVLIDSLPFTSHG-----	526
ACSL4	KGVEGTWVDICNNPAMEAEILKEIREAANAMKLERFEIPIKVRLSPWPPTGLVTD AF	648
Motif V		
ACSS3	KIPRSALS AIVNGKPYKITSTIE--DPSIFGHVEEMLKQA	686
ACSM4	KIKRNVL-----RDQEWGR-----	580
ACSF2	KIQKFKL-----REQMERHLNL-----	615
ACSF3	KIDKKAL-----IRHFHPS-----	576
ACSF4	KIDVSELNKIYLN YINL KSENKLSGKEDLWEKLQYLWKST (566tr)	
ACSL4	KLKRKEL-----RNHYLKDIERMYGGK	670

Fig. 2.—Continued.

outgroup with less support. Also, the long-chain clade had 35 members (including the five human proteins ACSL1, ACSL3, ACSL4, ACSL5, and ACSL6). Of these proteins, a subgroup of 14 including ACSL3 and ACSL4 was evident. Nonetheless, all of these proteins share similar substrate specificity and thus were classified as a single subfamily. Finally, in addition to the 10 major clades, an outlier containing a single yeast protein (sc6) was noted. This protein, known as Pcs60p or Fatp2, is a peroxisomal protein predicted to have ACS activity (46). We previously reported that Fatp2 belonged to an ACS subfamily that contains fungal, plant, and bacterial, but no mammalian, enzymes (11).

#### Structure-function correlations and identification of additional conserved ACS domains

At present, our knowledge of structure-function relationships among the various ACSs remains limited, particularly with respect to the mammalian enzymes. Mutagenesis ex-

periments involving bacterial (38, 39, 47), yeast (48), and plant (49) ACSs and related proteins have identified several residues that are critical for enzyme activity. Crystal structures of yeast or bacterial acetyl-coenzyme A synthetases (ACSacs; members of the ACSS subfamily) from the bacterium *Salmonella enterica* (12) and the yeast *S. cerevisiae* (14), and a putative bacterial ACSL from the extreme thermophile *Thermus thermophilus* (ttACS; a member of the ACSL subfamily) (13), have recently been described, allowing further predictions of functional residues. Not surprisingly, many of the amino acids identified as critical for enzymatic activity are those found in either Motif I or Motif II, as these are the most highly conserved residues. We hypothesized that examination of the human ACS sequences should permit the identification of additional conserved residues or domains that may be important for substrate binding, catalysis, enzyme regulation, or protein-protein interactions.

TABLE 4. ACS activity of human ACSF2 and ACSF3

Substrate	<i>pcDNA3</i>	ACSF2	<i>pcDNA3</i> <sup>a</sup>	ACSF3 <sup>a</sup>
	<i>nmol/20 min/mg protein</i>			
C8:0	4.6 ± 4.4	53 ± 2 <sup>b</sup>	9.9 ± 0.6	12.1 ± 1.0
C16:0	32 ± 9	38 ± 3	40 ± 3	43 ± 3
C24:0	0.8 ± 0.2	1.0 ± 0.1	1.3 ± 0.1	2.3 ± 0.5 <sup>c</sup>

Full-length cDNAs encoding human ACSF2 and ACSF3 were cloned into the mammalian expression vector *pcDNA3*, as described in Materials and Methods. COS-1 cells were transfected by electroporation. After 3 days, cells were harvested, frozen/thawed, and assayed for their ability to activate a representative medium-chain (C8:0), long-chain (C16:0), or very long-chain (C24:0) fatty acid. Results are presented as means ± SD of three (ACSF3) or four (ACSF2) independent transfection experiments.

<sup>a</sup>For the measurement of ACSF3 activity (and corresponding vector control), Triton X-100 was included in assays at a final concentration of 0.1%.

<sup>b</sup>Statistically significant ( $P < 0.001$ ).

<sup>c</sup>Statistically significant ( $P < 0.05$ ).

Hisanaga et al. (13) defined four structurally significant domains in ttACS that they referred to as the P-loop and the L-, A-, and G-motifs. The P-loop residues are the last 9 of 10 amino acids that constitute Motif I (Table 1), and the L-motif consists of 6 amino acids (432-DRLKDL-437) found within Motif II (T416 through E451) of ttACS (13). The A-motif of ttACS is a sequence of seven amino acids (323-GYGLTET-329) located between Motifs I and II. Examination of the human ACSs revealed that 22 of the 26 proteins contained a related sequence (consensus, YGXTTE), herein referred to as Motif III (Table 1, Fig. 1C). Motif III was found 70–100 residues upstream of Motif II. Only the three members of the ACSS family and ACSF1 lacked Motif III. However, the related sequence [W,F]WQTE was found in a similar region of the ACSS proteins but not in ACSF1. The WWQTE motif was also found instead of the A-motif in *S. enterica* ACSAc (12). A nine amino acid G- (or gate) motif (226-VPMFHVNAW-234) is located just downstream of Motif I in ttACS (13). A conserved motif (Motif IV) homologous to the first five residues of the gate motif, with consensus LPLXH, was found in 15 human ACSs, including all ACSL, ACSVL, and ACSBG family members, as well as ACSF2 and ACSF3 (Table 1, Fig. 1D).

Mutation of a lysine residue (K592) near the C terminus of *S. enterica* propionyl-CoA synthetase (a member of the ACSS subfamily) to either alanine or glutamate prevented the conversion of propionate to propionyl-CoA (47). Interestingly, K592 was found to be essential for the formation of propionyl-AMP but not for the conversion of propionyl-AMP to propionyl-CoA. This residue corresponds to K609 of *S. enterica* ACSAc, which is regulated by acetylation/deacetylation (50). Acetylation effectively blocks the formation of acetyl-AMP (reaction I shown above) without affecting thioesterification to CoA (reaction II) (50). Deacetylation, catalyzed by the Sir2 protein in a NAD-dependent reaction, activates the enzyme by releasing this inhibition of reaction I (50). The motif containing the conserved lysine (underlined) in both *S. enterica* enzymes is PKTRSGKXXR (50). This motif, with consensus PKTX[S,T]GKIX[R,K], can be found in 13 human ACS

sequences, including all ACSS and ACSM family members and ACSF1, ACSF2, ACSF3, and ACSF4 (Motif V) (Table 1, Fig. 1E). Although the sequence preceding the conserved K in the consensus motif is not found in members of the ACSL, ACSVL, or ACSBG families, the KXX [R,K] motif is present (Table 1). The KXX[R,K] motif is found in 24 of the human ACS sequences; in ACSF1 it is KXXE, and in ACSF4 it is KXXV.

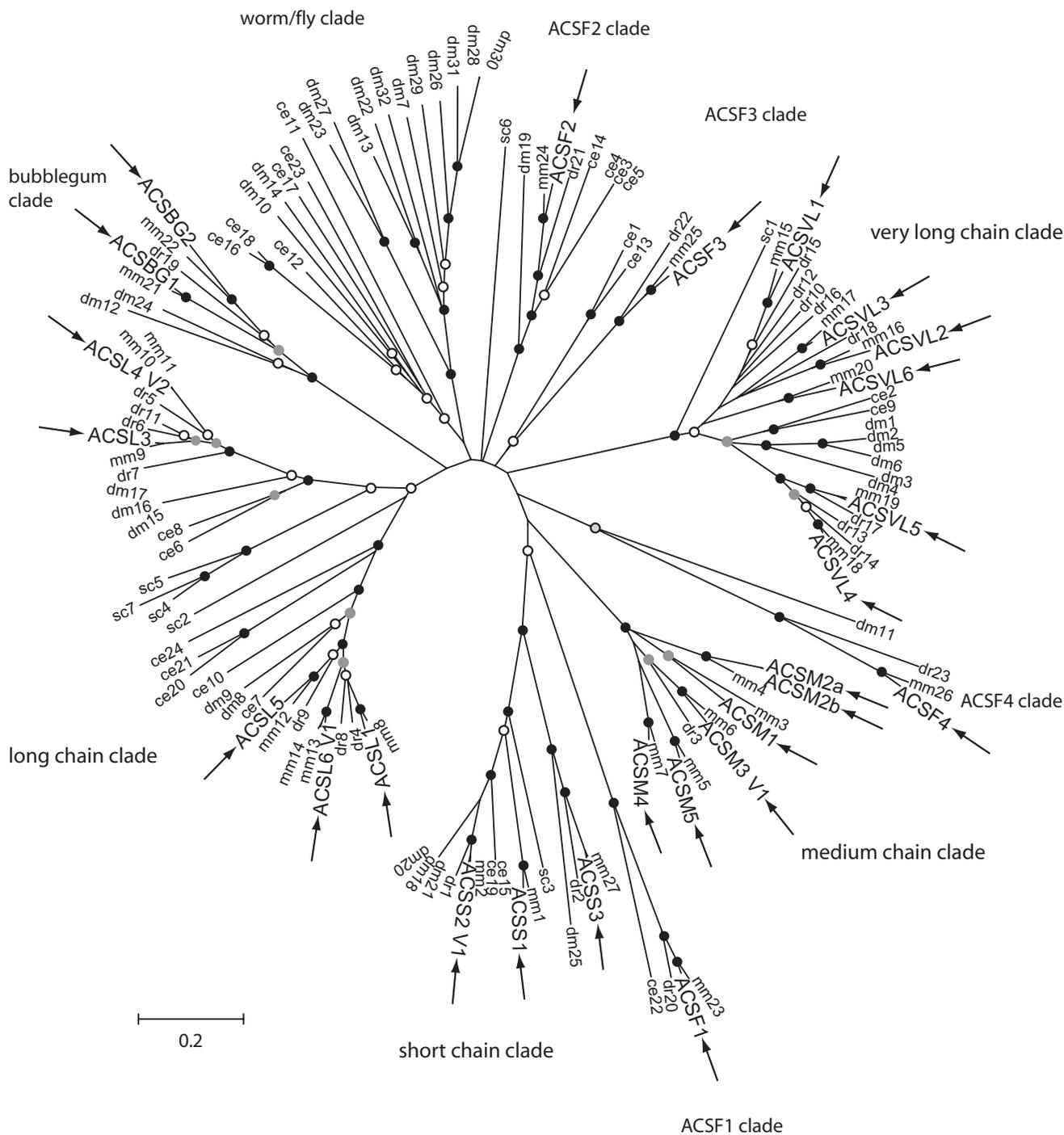
## DISCUSSION

In this report, we describe what we believe to be the full complement of ACS genes in the human genome. Our investigations support the conclusion that there are 26 distinct ACS genes. All of these genes likely encode proteins, because *a*) 20 of them have been shown previously to encode ACSs in humans or other mammals, *b*) cDNAs derived from two additional gene sequences encode enzymatically active ACSs when expressed in COS cells, and *c*) multiple ESTs exist for each sequence, suggesting the existence of RNA transcripts.

The strategy used here to identify human ACS sequences using two conserved amino acid sequence motifs built upon previous work by us (11) and others (37–39, 49, 51–53). All enzymatically active ACSs, including ACSF2 and ACSF3 described in this work, contained both Motifs I and II. The locations of Motifs I and II, and the distance separating these sequences, were similar in all of these sequences, and most human ACSs were similar in size. Two exceptions were ACSM3\_v2 and ACSF4 (Table 3). Transcript variant 2 of ACSM3 was shorter than other ACSs (438 amino acids) and lacked Motif II. Although experimental proof is not yet available, we predict that ACSM3\_v2 is devoid of ACS activity, although it may have other biological functions.

ACSF4 was identified as an ACS gene in this report. Although the ACSF4 amino acid sequence contains structural features suggesting that it belongs to the ACS family (e.g., the relative positions of Motifs I and II with respect to the initiator methionine and with respect to each other), the protein is substantially larger than all other ACSs (1,098 amino acids) and had previously been identified as 2-amino adipic 6-semialdehyde dehydrogenase (43). ACSF4 is homologous to the yeast enzyme LYS2, which is required for lysine biosynthesis in lower eukaryotes (54). In humans, in which lysine is an essential amino acid, this enzyme operates in the reverse direction and serves a catabolic function. In *S. cerevisiae*, LYS2 is activated by the phosphopantetheinylation of serine 880 in an ATP-dependent reaction catalyzed by LYS5; CoA is the donor of the phosphopantetheine group (55). A homologous residue, serine 589, is found in ACSF4. Thus, it is possible that conserved ACS motifs in ACSF4 serve a function other than fatty acid activation. Further experimentation is necessary to establish whether ACSF4 has ACS activity.

The arginine residue found in the center of Motif II (Table 1) is essentially invariant and is present in all ACSs from all species from archaea to humans (P. A. Watkins,



**Fig. 3.** Phylogenetic analysis of human ACSs. Amino acid sequences of 26 human and 113 other ACSs (from *M. musculus*, *D. rerio*, *D. melanogaster*, *C. elegans*, and *S. cerevisiae*) were aligned using MUSCLE, as described in Materials and Methods. The human proteins are indicated by arrows and larger font size. The evolutionary relationships were inferred using the neighbor-joining method (34) as implemented in MEGA4 software (33). Branch lengths are indicated (bar = 0.2 amino acid substitutions per site). Bootstrap analysis was used to assess the reliability of specific clades in the tree, as described in Materials and Methods. Black circles,  $\geq 95\%$  bootstrap support; gray circles, 80–95%; white circles, 50–80%. These results support the assignment of ACSs into 10 major clades, shown in groups with labels. Accession numbers for the human sequences are listed in Table 3, and accession numbers for nonhuman sequences are listed in supplementary Table III.

unpublished observation). To the best of our knowledge, only one bona fide ACS, human ACSBG2, has a different residue (histidine) at this position. ACSBG2 sequences from chimpanzees, rhesus monkeys, dogs, rats, and mice all retain the conserved arginine. We previously reported that the

arginine-to-histidine substitution in human ACSBG2 decreased the pH optimum of the enzyme (25), but the physiologic significance of this change remains obscure.

There was a reasonably good correlation between the phylogenetic placement of ACS sequences into families

of structurally related proteins and the substrate preferences of these enzymes. Enzymes known to activate short-, medium-, long-, and very long-chain FAs were assigned to the ACSS, ACSM, ACSL, and ACSVL families, respectively. For a few enzymes, particularly ACSS3, ACSM5, and ACSM2A, the appropriateness of their placement remains speculative until confirmed experimentally. The two ACSBG family members have unique substrate specificities. Although ACSBG1 was thought to activate VLCFA substrates based on overexpression studies (11), subsequent investigation of the endogenous enzyme using RNA interference revealed a high specificity for palmitic acid (C16:0) (56). ACSBG2 preferentially activates oleic (C18:1) and linoleic (C18:2) acid substrates (25).

In a previous study, we identified an ACS subfamily that contained Fat2p, a putative ACS from *S. cerevisiae* (46). This family contained proteins from *Schizosaccharomyces pombe*, *Mycobacterium tuberculosis*, and *Arabidopsis thaliana*, but no human or other mammalian homologs were identified (11). Interestingly, Fat2p (designated sc6 and located between the worm/fly clade and the ACSF2 clade in Fig. 3) also appears to have no zebrafish, fruitfly, or worm homologs. Enzymatic activity of Fatp2 has not yet been verified.

Knowing the amino acid sequences of all human ACSs facilitated the evaluation of conserved domains. This knowledge should enhance our understanding of structure-function relationships in these enzymes. Motif I (Table 1, Fig. 1A) includes the P-loop described by Hisanaga et al. (13). Often referred to as the AMP binding domain, the P-loop is found in close proximity to the adenosine moiety and helps maintain the substrate in the proper orientation. Mutagenesis of several residues within Motif I in the *Escherichia coli* ACSL, FadD, resulted in decreased enzyme activity (38). Mutation of the first Motif I residue in FadD, Y213, nearly abolished activity, whereas mutations in residues 2, 4, 5, and 10 (T214, G216, T217, and K222) led to reduced catalytic efficiency. A similar result was found when the homologous lysine (K248) of *S. enterica* propionyl-CoA synthetase, a member of the ACSS family, was mutated (47). Mutagenesis of Motif I residues 1 and 5 (Y256 and T260) of *S. cerevisiae* Fat1p, a member of the ACSVL family, produced only a mild reduction in enzyme activity, whereas a mutation in residue 3 (S258) had a more severe effect on activity (48).

The Motif II sequence (Table 1, Fig. 1B) contains the L-motif (432-DRLKDL-437) that in ttACS acts as a linker between the large N-terminal domain and the smaller C-terminal domain (13). The linker region is thought to be critical for catalytic function, as it facilitates a conformational change upon ATP binding that permits subsequent binding of the fatty acyl and/or CoA substrates. This hypothesis was reinforced by examination of the crystal structures of ACSAc from both yeast and bacteria, in which the "hinge" residue was identified as an aspartate residue (corresponding to the underlined D in the L-motif) (12, 14). This aspartate residue is conserved in 18 human ACS sequences. Because of the variability at this position in the other eight sequences (E, E, H, H, G, N, S, and V), this residue was not included in the Motif II consensus. However, if the conformational change predicted by the three

available crystal structures is applicable to all ACSs, a hinge amino acid is likely to be critical for enzyme activity.

The signature motif, identified by Black and coworkers (39) in a group of enzymes from diverse species belonging to the ACSL family, overlaps with the first 20 residues of Motif II and contains the first four amino acids of the linker motif. Mutagenesis of several Motif II residues in FadD, including amino acids 1 and 3 (T436 and D438), and the highly conserved arginine (R453) significantly decreased catalytic function (39). Similarly, mutagenesis of the corresponding aspartate (D508) and arginine (R523) of Fat1p was deleterious for fatty acid activation (48). Interestingly, mutations in two Fat1p Motif II residues (Y519 and S536) that are not well conserved between the different ACS families and thus are not part of the consensus also decreased catalytic function. A lysine found six residues downstream of the conserved arginine was critical for the activity of the related plant enzyme, coumarate-CoA ligase (49), and was proposed to participate catalytically in ttACS (13). Although this lysine can be found in the five human ACSL proteins and in ACSF4, it is not conserved among the other 20 human ACSs.

Motif III (Table 1, Fig. 1C) was found in nearly all human ACSs and is part of the A- (or adenine) motif of ttACS (13). This region has been described as an ATP/AMP binding domain in other ACSs (38, 47, 57). Structural analysis of ttACS showed that Y324 was an adenine binding residue (13). Site-directed mutagenesis of the glutamate residue of Motif III in the *E. coli* ACSL, FadD (E361), abolished enzyme activity (38). The crystal structure of *S. enterica* ACSAc revealed that the conserved glutamate residue of Motif III is positioned near oxygen O1 of the AMP phosphate (12). The tryptophan residues in this loop, like the adenine binding Y324 of ttACS, are in proximity to the adenine ring, suggesting an essential role in substrate binding or stabilization.

Motif IV (Table 1, Fig. 1D) was found in 15 human ACS sequences. This motif comprises the first five residues of the nine amino acid G- (or gate) motif (226-VPMFHVNAW-234) of ttACS (13). From the crystal structure of ttACS, it was proposed that the indole ring of W234 acts as a gate and blocks the entry of fatty acids into its substrate binding tunnel unless ATP is first bound, resulting in a conformational change that swings the gate open (13). However, a tryptophan residue corresponding to W234 was not found in any human ACS sequences. In contrast, although no highly conserved sequences were identified, candidate gate tryptophan residues were found in the expected region of ACSS and ACSM family members as well as in ACSF1. Because ACSL, ACSVL, and ACSBG enzymes activate longer chain fatty acid substrates, a corresponding gate residue may be located elsewhere in the structure.

The conserved lysine residue of Motif V (Table 1, Fig. 1E), required for the catalytic activity of *S. enterica* propionyl-CoA synthetase (47), was recently found to be essential for the ACS activity of murine ACSF2 (D. Maignel and P. A. Watkins, unpublished). In the yeast ACSAc crystallized as a binary complex with AMP, the corresponding lysine residue was located near the catalytic site (14). In contrast, this amino

acid was found on the surface of the bacterial ACSAc crystallized as a ternary complex with propyl-AMP and CoA (12). These observations are consistent with the proposed role of this residue in the first half-reaction and with the subsequent large conformational change (rotation of  $\sim 140^\circ$ ) in the C-terminal domain that may help create the CoA binding pocket. In ttACS crystallized as a complex with myristoyl-AMP, the homologous 524-KXXK-527 motif is part of a loop-helix also found on the surface of the protein near the C terminus (13). However, K527 and not K524 was one of three residues proposed to stabilize the closed conformation of the protein (after ATP binding) by forming noncovalent interactions with residues of the L-motif (found within motif II) and the N-terminal domain (13). Evidence for the control of mammalian ACSS1 and ACSS2 activity by reversible acetylation of the conserved Motif V lysine residue was published recently, solidifying the importance of this domain (58, 59).

For several human ACS genes, alternative transcripts were identified. To date, only the two variants of ACSL6 have been investigated at the biochemical level. These variants differ in a 27 amino acid stretch that encompasses Motif IV and the gate domain. These two variants differed primarily in their  $K_m$  for ATP (45). Soupene and Kuypers (60) recently reported additional transcript variants of human ACSL family members, which were isolated primarily from erythroid cells using PCR. These authors found two additional variants of ACSL1 and three additional variants of ACSL6. We did not include these variants in Table 3, either because *a*) full-length sequences were not available or *b*) we could not find corroborating evidence in public databases to support the existence of these variants. One ACSL1 variant (a 373 amino acid fragment) contained an alternatively spliced exon encompassing Motif IV, highly similar to the situation with the ACSL6 variants. Although no supporting sequences were present in the nonredundant databases, one relevant human EST (from trachea) was found, suggesting that this represents a true variant. A fragment of a third ACSL1 variant (93 amino acids) was also reported, but no nonredundant or EST sequences were found to substantiate it. Two of the additional ACSL6 variants, designated ACSL6\_v3 (622 amino acids) and ACSL6\_v5 (712 amino acids), could represent full-length ACS sequences, but the third variant, ACSL6\_v4, was a fragment of 115 amino acids. However, we were unable to detect unequivocal supporting evidence for any of these ACSL6 variants in nonredundant or EST databases. Further studies are thus needed to establish the validity of these ACSL variants.

Finally, to establish the validity of our ACS identification strategy, we demonstrated that two candidate human ACSs that had not been studied previously are indeed enzymatically active. For this, we chose two of the four ACS sequences that were evolutionarily divergent from established ACS families, namely ACSF2 and ACSF3. Because these putative ACSs were "orphans," we had no preconceived notions regarding their substrate chain length preferences. One of these, ACSF2, preferred a medium-chain substrate, whereas the other, ACSF3, preferred the very

long-chain substrate. More studies of these enzymes are needed to establish their complete substrate profiles and their normal role(s) in lipid metabolism. ■

The authors thank Dr. Joseph Hacia (University of Southern California) for invaluable advice and guidance on bioinformatics techniques and programs. The authors also thank Hanh Ho, Stefanie Alzate, Shawn Zardouz, Erica Aronson, and Alicia Rizzo for excellent technical and bioinformatics assistance. We are grateful to Dr. Varsha Khodiyar, gene nomenclature advisor for the HUGO Gene Nomenclature Committee, for her insightful thoughts and suggestions regarding ACS nomenclature. This work was supported by National Institutes of Health Grants HD-10981, HD-24061, and NS-37355.

## REFERENCES

1. Watkins, P. A. 1997. Fatty acid activation. *Prog. Lipid Res.* **36**: 55–83.
2. Webster, L. T., Jr. 1965. Studies of the acetyl coenzyme A synthetase reaction. II. Crystalline acetyl coenzyme A synthetase. *J. Biol. Chem.* **240**: 4158–4163.
3. Webster, L. T., Jr., L. D. Gerowin, and L. Rakitak. 1965. Purification and characteristics of a butyryl coenzyme A synthetase from bovine heart mitochondria. *J. Biol. Chem.* **240**: 29–33.
4. Mahler, H. R., S. J. Wakil, and R. M. Bock. 1953. Studies on fatty acid oxidation. I. Enzymatic activation of fatty acids. *J. Biol. Chem.* **204**: 453–468.
5. Kornberg, A., and W. E. Pricer, Jr. 1953. Enzymatic synthesis of the coenzyme A derivatives of long chain fatty acids. *J. Biol. Chem.* **204**: 329–343.
6. Singh, I., M. S. Kang, and L. A. Phillips. 1982. Lignoceroyl:CoA ligase activity in rat brain microsomal fraction. *Fed. Proc.* **41**: 1192.
7. Tanaka, T., K. Hosaka, M. Hoshimaru, and S. Numa. 1979. Purification and properties of long-chain acyl-coenzyme-A synthetase from rat liver. *Eur. J. Biochem.* **98**: 165–172.
8. Hall, A. M., A. J. Smith, and D. A. Bernlohr. 2003. Characterization of the acyl CoA synthetase activity of purified murine fatty acid transport protein 1. *J. Biol. Chem.* **278**: 43008–43013.
9. Hall, A. M., B. M. Wiczer, T. Herrmann, W. Stremmel, and D. A. Bernlohr. 2005. Enzymatic properties of purified murine fatty acid transport protein 4 and analysis of acyl-CoA synthetase activities in tissues from FATP4 null mice. *J. Biol. Chem.* **280**: 11948–11954.
10. Steinberg, S. J., S. J. Wang, D. G. Kim, S. J. Mihalik, and P. A. Watkins. 1999. Human very-long-chain acyl-CoA synthetase: cloning, topography, and relevance to branched-chain fatty acid metabolism. *Biochem. Biophys. Res. Commun.* **257**: 615–621.
11. Steinberg, S. J., J. Morgenthaler, A. K. Heinzer, K. D. Smith, and P. A. Watkins. 2000. Very long-chain acyl-CoA synthetases. Human "bubblegum" represents a new family of proteins capable of activating very long-chain fatty acids. *J. Biol. Chem.* **275**: 35162–35169.
12. Gulick, A. M., V. J. Starai, A. R. Horswill, K. M. Homick, and J. C. Escalante-Semerena. 2003. The 1.75 Å crystal structure of acetyl-CoA synthetase bound to adenosine-5'-propylphosphate and coenzyme A. *Biochemistry.* **42**: 2866–2873.
13. Hisanaga, Y., H. Ago, N. Nakagawa, K. Hamada, K. Ida, M. Yamamoto, T. Hori, Y. Arai, M. Sugahara, S. Kuramitsu, et al. 2004. Structural basis of the substrate specific two-step catalysis of long chain fatty acyl-CoA synthetase dimer. *J. Biol. Chem.* **279**: 31717–31726.
14. Jogl, G., and L. Tong. 2004. Crystal structure of yeast acetyl-coenzyme A synthetase in complex with AMP. *Biochemistry.* **43**: 1425–1431.
15. Katoh, K., K. Kuma, H. Toh, and T. Miyata. 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res.* **33**: 511–518.
16. Edgar, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**: 1792–1797.
17. Schneider, T. D., and R. M. Stephens. 1990. Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res.* **18**: 6097–6100.
18. Crooks, G. E., G. Hon, J. M. Chandonia, and S. E. Brenner. 2004. WebLogo: a sequence logo generator. *Genome Res.* **14**: 1188–1190.

19. Watkins, P. A., E. V. Ferrell, Jr., J. I. Pedersen, and G. Hoefler. 1991. Peroxisomal fatty acid beta-oxidation in HepG2 cells. *Arch. Biochem. Biophys.* **289**: 329–336.
20. Dole, V. P. 1956. A relation between non-esterified fatty acids in plasma and the metabolism of glucose. *J. Clin. Invest.* **35**: 150–154.
21. Mashek, D. G., K. E. Bornfeldt, R. A. Coleman, J. Berger, D. A. Bernlohr, P. Black, C. C. DiRusso, S. A. Farber, W. Guo, N. Hashimoto, et al. 2004. Revised nomenclature for the long chain mammalian acyl-CoA synthetase gene family. *J. Lipid Res.* **45**: 1958–1961.
22. Fujino, T., Y. A. Takei, H. Sone, R. X. Ioka, A. Kamataki, K. Magoori, S. Takahashi, J. Sakai, and T. T. Yamamoto. 2001. Molecular identification and characterization of two medium-chain acyl-CoA synthetases, MACS1 and the Sa gene product. *J. Biol. Chem.* **276**: 35961–35966.
23. Vessey, D. A., E. Lau, M. Kelley, and R. S. Warren. 2003. Isolation, sequencing, and expression of a cDNA for the HXMA form of xenobiotic/medium-chain fatty acid:CoA ligase from human liver mitochondria. *J. Biochem. Mol. Toxicol.* **17**: 1–6.
24. Min, K. T., and S. Benzer. 1999. Preventing neurodegeneration in the *Drosophila* mutant bubblegum. *Science*. **284**: 1985–1988.
25. Pei, Z., Z. Jia, and P. A. Watkins. 2006. The second member of the human and murine bubblegum family is a testis- and brainstem-specific acyl-CoA synthetase. *J. Biol. Chem.* **281**: 6632–6641.
26. Stahl, A., R. E. Gimeno, L. A. Tartaglia, and H. F. Lodish. 2001. Fatty acid transport proteins: a current view of a growing family. *Trends Endocrinol. Metab.* **12**: 266–273.
27. Hirsch, D., A. Stahl, and H. F. Lodish. 1998. A family of fatty acid transporters conserved from *Mycobacterium* to man. *Proc. Natl. Acad. Sci. USA.* **95**: 8625–8629.
28. Uchida, Y., N. Kondo, T. Orii, and T. Hashimoto. 1996. Purification and properties of rat liver peroxisomal very-long-chain acyl-CoA synthetase. *J. Biochem. (Tokyo)*. **119**: 565–571.
29. Uchiyama, A., T. Aoyama, K. Kamijo, Y. Uchida, N. Kondo, T. Orii, and T. Hashimoto. 1996. Molecular cloning of cDNA encoding rat very long-chain acyl-CoA synthetase. *J. Biol. Chem.* **271**: 30360–30365.
30. Steinberg, S. J., S. J. Mihalik, D. G. Kim, D. A. Cuebas, and P. A. Watkins. 2000. The human liver-specific homolog of very long-chain acyl-CoA synthetase is cholate:CoA ligase. *J. Biol. Chem.* **275**: 15605–15608.
31. Mihalik, S. J., S. J. Steinberg, Z. Pei, J. Park, D. G. Kim, A. K. Heinzer, G. Dacremont, R. J. Wanders, D. A. Cuebas, K. D. Smith, et al. 2002. Participation of two members of the very long-chain acyl-CoA synthetase family in bile acid synthesis and recycling. *J. Biol. Chem.* **277**: 24771–24779.
32. Swofford, D. L. 2003. PAUP\*: Phylogenetic Analysis Using Parsimony (\*and Other Methods). Sinauer Associates, Sunderland, MA.
33. Tamura, K., J. Dudley, M. Nei, and S. Kumar. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* **24**: 1596–1599.
34. Saitou, N., and M. Nei. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**: 406–425.
35. Berg, P. 1956. Acyl adenylates: an enzymatic mechanism of acetate activation. *J. Biol. Chem.* **222**: 991–1013.
36. Jencks, W. P., and F. Lipmann. 1957. Studies on the initial step of fatty acid activation. *J. Biol. Chem.* **225**: 207–223.
37. Karan, D., M. Lesbats, J. R. David, and P. Capy. 2003. Evolution of the AMP-forming acetyl-CoA synthetase gene in the *Drosophilidae* family. *J. Mol. Evol.* **57** (Suppl. 1): 297–303.
38. Weimar, J. D., C. C. DiRusso, R. Delio, and P. N. Black. 2002. Functional role of fatty acyl coenzyme A synthetase in the transmembrane movement and activation of exogenous long-chain fatty acids: amino acid residues within the ATP/AMP signature motif of FadD of *Escherichia coli* are required for enzyme activity and fatty acid transport. *J. Biol. Chem.* **277**: 29369–29376.
39. Black, P. N., Q. Zhang, J. D. Weimar, and C. C. DiRusso. 1997. Mutational analysis of a fatty acyl-coenzyme A synthetase signature motif identifies seven amino acid residues that modulate fatty acid substrate specificity. *J. Biol. Chem.* **272**: 4896–4903.
40. Iwai, N., T. Mannami, H. Tomoike, K. Ono, and Y. Iwanaga. 2003. An acyl-CoA synthetase gene family in chromosome 16p12 may contribute to multiple risk factors. *Hypertension*. **41**: 1041–1046.
41. Iwai, N., T. Katsuya, T. Mannami, J. Higaki, T. Ogiwara, K. Kokame, J. Ogata, and S. Baba. 2002. Association between SAH, an acyl-CoA synthetase gene, and hypertriglyceridemia, obesity, and hypertension. *Circulation*. **105**: 41–47.
42. Oka, Y., K. Kobayakawa, H. Nishizumi, K. Miyamichi, S. Hirose, A. Tsuboi, and H. Sakano. 2003. O-MACS, a novel member of the medium-chain acyl-CoA synthetase family, specifically expressed in the olfactory epithelium in a zone-specific manner. *Eur. J. Biochem.* **270**: 1995–2004.
43. Wang, L., C. Jil, Y. Xu, J. Xu, J. Dai, Q. Wu, M. Wu, X. Zou, L. Sun, S. Gu, et al. 2005. Cloning and characterization of a novel human homolog\* of mouse U26, a putative PQQ-dependent AAS dehydrogenase. *Mol. Biol. Rep.* **32**: 47–53.
44. Cao, Y., E. Traer, G. A. Zimmerman, T. M. McIntyre, and S. M. Prescott. 1998. Cloning, expression, and chromosomal localization of human long-chain fatty acid-CoA ligase 4 (FACL4). *Genomics*. **49**: 327–330.
45. Van Horn, C. G., J. M. Caviglia, L. O. Li, S. Wang, D. A. Granger, and R. A. Coleman. 2005. Characterization of recombinant long-chain rat acyl-CoA synthetase isoforms 3 and 6: identification of a novel variant of isoform 6. *Biochemistry*. **44**: 1635–1642.
46. Blobel, F., and R. Erdmann. 1996. Identification of a yeast peroxisomal member of the family of AMP-binding proteins. *Eur. J. Biochem.* **240**: 468–476.
47. Horswill, A. R., and J. C. Escalante-Semerena. 2002. Characterization of the propionyl-CoA synthetase (PrpE) enzyme of *Salmonella enterica*: residue Lys592 is required for propionyl-AMP synthesis. *Biochemistry*. **41**: 2379–2387.
48. Zou, Z., C. C. DiRusso, V. Ctrnacta, and P. N. Black. 2002. Fatty acid transport in *Saccharomyces cerevisiae*. Directed mutagenesis of FAT1 distinguishes the biochemical activities associated with Fat1p. *J. Biol. Chem.* **277**: 31062–31071.
49. Stuible, H., D. Buttner, J. Ehling, K. Hahlbrock, and E. Kombrink. 2000. Mutational analysis of 4-coumarate:CoA ligase identifies functionally important amino acids and verifies its close relationship to other adenylate-forming enzymes. *FEBS Lett.* **467**: 117–122.
50. Starai, V. J., I. Celic, R. N. Cole, J. D. Boeke, and J. C. Escalante-Semerena. 2002. Sir2-dependent activation of acetyl-CoA synthetase by deacetylation of active lysine. *Science*. **298**: 2390–2392.
51. Babbitt, P. C., G. L. Kenyon, B. M. Martin, H. Charest, M. Slyvestre, J. D. Scholten, K. H. Chang, P. H. Liang, and D. Dunaway-Mariano. 1992. 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*. **31**: 5594–5604.
52. Marahiel, M. A., T. Stachelhaus, and H. D. Mootz. 1997. Modular peptide synthetases involved in nonribosomal peptide synthesis. *Chem. Rev.* **97**: 2651–2674.
53. Schroder, J. 1989. Protein sequence homology between plant 4-coumarate:CoA ligase and firefly luciferase. *Nucleic Acids Res.* **17**: 460.
54. Bhattacharjee, J. K. 1985.  $\alpha$ -Amino adipate pathway for the biosynthesis of lysine in lower eukaryotes. *Crit. Rev. Microbiol.* **12**: 131–151.
55. Ehmann, D. E., A. M. Gehring, and C. T. Walsh. 1999. Lysine biosynthesis in *Saccharomyces cerevisiae*: mechanism of alpha-amino adipate reductase (Lys2) involves posttranslational phosphopantetheinylation by Lys5. *Biochemistry*. **38**: 6171–6177.
56. Pei, Z., N. A. Oey, M. M. Zuidervaart, Z. Jia, Y. Li, S. J. Steinberg, K. D. Smith, and P. A. Watkins. 2003. The acyl-CoA synthetase “bubblegum” (lipidosin): further characterization and role in neuronal fatty acid beta-oxidation. *J. Biol. Chem.* **278**: 47070–47078.
57. Chang, K. H., H. Xiang, and D. Dunaway-Mariano. 1997. Acyl-adenylate motif of the acyl-adenylate/thioester-forming enzyme superfamily: a site-directed mutagenesis study with the *Pseudomonas* sp. strain CBS3 4-chlorobenzoate:coenzyme A ligase. *Biochemistry*. **36**: 15650–15659.
58. Schwer, B., J. Bunkenborg, R. O. Verdin, J. S. Andersen, and E. Verdin. 2006. Reversible lysine acetylation controls the activity of the mitochondrial enzyme acetyl-CoA synthetase 2. *Proc. Natl. Acad. Sci. USA.* **103**: 10224–10229.
59. Hallows, W. C., S. Lee, and J. M. Denu. 2006. Sirtuins deacetylate and activate mammalian acetyl-CoA synthetases. *Proc. Natl. Acad. Sci. USA.* **103**: 10230–10235.
60. Soupene, E., and F. A. Kuypers. 2006. Multiple erythroid isoforms of human long-chain acyl-CoA synthetases are produced by switch of the fatty acid gate domains. *BMC Mol. Biol.* **7**: 21.
61. Watkins, P. A., J. F. Lu, S. J. Steinberg, S. J. Gould, K. D. Smith, and L. T. Braiterman. 1998. Disruption of the *Saccharomyces cerevisiae* FAT1 gene decreases very long-chain fatty acyl-CoA synthetase activity and elevates intracellular very long-chain fatty acid concentrations. *J. Biol. Chem.* **273**: 18210–18219.
62. Watkins, P. A., J. Pevsner, and S. J. Steinberg. 1999. Human very long-chain acyl-CoA synthetase and two human homologs: initial characterization and relationship to fatty acid transport protein. *Prostaglandins Leukot. Essent. Fatty Acids.* **60**: 323–328.