Evidence for 26 distinct acyl-coenzyme A synthetase genes in the human genome

Paul A. Watkins, Dony Maiguel, Zhenzhou Jia, and Jonathan Pevsner
Kennedy Krieger Institute and Departments of Neurology and Neuroscience, 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, Paul A., Dony Maiguel, Zhenzhou Jia, and Jonathan 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

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

Supplemental Material can be found at:
http://www.jlr.org/cgi/content/full/M700378-JLR200/DC1

Supplemental Table 1: Fatty acid chain length and preference

Supplemental Table 2: Fatty acid desaturation by ACS

Supplemental Table 3: Functional properties of ACSs

Abbreviations: ACS, acyl-coenzyme A synthetase; ACSSc, yeast or bacterial acyl-coenzyme A synthetase; ACBG, bubblegum ACS; ACSF, ACS family; ACSL, long-chain ACS; ACSM, medium-chain ACS; ACSS, short-chain ACS; ACSV, 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.

1To whom correspondence should be addressed.
e-mail: watkins@kennedykrieger.org

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

Copyright © 2007 by the American Society for Biochemistry and Molecular Biology, Inc.
This article is available online at http://www.jlr.org
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 search the human nonredundant protein database. These searches yielded no additional candidate ACS genes present. With the exception of one protein tentatively identified as aminoadipic 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 ACSF3 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'-attgtgctcgagctgtagctg-3', which incorporates a BamHI restriction site (underlined), and reverse oligonucleotide 5'-tgtccgaccagcagcag-3', which incorporates a NotI restriction site (underlined). After digestion with restriction enzymes, the resulting 1,847 bp fragment was cloned into the BamHI and NotI 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'-ccgcaattgattaagattg-3', which incorporates an EcoRI site (underlined), and reverse oligonucleotide 5'-ggatccgacccgctggaccttctg-3', which incorporates a XhoI site (underlined). After restriction enzyme digestion of the PCR product, the 1,865 bp fragment was cloned into the EcoRI and XhoI 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 [1-14C]octanoic acid (American Radiolabeled Chemicals). [1-14C]palmitic acid (Moravek Biochemicals), or [1-14C]lignoceric acid (Moravek Biochemicals) as described previously (19). Final fatty acid concentrations in assays were 400 μM for octanoate and 20 μM for palmitate and lignocerate and included ~100,000 dpm (1 nmol) of labeled fatty acid. Fatty acids were solubilized using α-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 MgCl2, 0.2 mM CoA, 0.2 mM DTT, and cell suspension (15 μg protein/assay for octanoate and palmitate, 60 μg protein/assay for lignocerate) in a total volume of 250 μl. 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 ACS, ACSM, and ACSBG subfamilies.

The rationale for the ACS family nomenclature is as follows. The older approved gene symbols for the two known human acetyl-CoA synthetases were ACS2 and ACS2L (for ACS2-like); there was no ACS1. Converting to the uniform nomenclature system for the ACS proteins, ACS2 became ACS2L and ACS2L became ACS1. The proposed name for a third subfamily member, identified in this work, is ACS3. The ACS family 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 ACS, 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 YTSGTTGLPK and FTSGTTGLPK 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 nonhuman 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 13 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). 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):

\[
\text{fatty acid + ATP} \rightarrow \text{fatty acyl-AMP + Pi} \quad (I)
\]

\[
\text{fatty acyl-AMP + CoA-SH} \rightarrow \text{fatty acyl-S-CoA + AMP} \quad (II)
\]
The first step is an ATP-dependent adenylation of the substrate with concomitant release of pyrophosphate (reaction I). CoA-SH then displaces the adenylate 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), is found within the 11 amino acid putative AMP binding domain signature in the PROSITE database (PDDoc00427, 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 ACSLs 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:

\[
T\text{-}G\text{-}D\text{-}X(7)\text{-}G\text{-}X(3)\text{-}[h]X(2)\text{-}R\text{-}X(4)\text{-}[h]X(3,4)\text{-}G\text{-}X(2)\text{-}[h]X(4)\text{-}[h]E
\]

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 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 ACSSs, ACSSMs, ACSLs, 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 ACSSM2A, ACSS3, ACSSM5, ACSSF2, ACSSF3, and ACSSF4 (Tables 2, 3). [Concurrent with this work, we established that murine ACSSF2 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, ACSSM2A and ACSSM5, have been published because of their homology to ACSSM3 (formerly known as SAH) (40, 41). However, the biological functions of ACSSM2A and ACSSM5 have not yet been characterized in any species.

The amino acid sequences of three candidate human ACSs (ACSS3, ACSSF2, and ACSSF3) have not been published previously. The sequence of rodent ACSSM4 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 ACSSL, ACSSL4, is included for comparison. Another of the candidate ACS genes, ACSSF4, was reported to encode 2-
<table>
<thead>
<tr>
<th>ACS</th>
<th>Motif I</th>
<th>Motif II</th>
<th>Motif III</th>
<th>Motif IV</th>
<th>Motif V</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short-chain ACS family</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACSS1</td>
<td>293</td>
<td>293</td>
<td>550</td>
<td>550</td>
<td>550</td>
</tr>
<tr>
<td>ACSS2</td>
<td>315</td>
<td>315</td>
<td>550</td>
<td>550</td>
<td>550</td>
</tr>
<tr>
<td>ACSS3</td>
<td>298</td>
<td>298</td>
<td>550</td>
<td>550</td>
<td>550</td>
</tr>
<tr>
<td>Medium-chain ACS family</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACMS1</td>
<td>225</td>
<td>225</td>
<td>550</td>
<td>550</td>
<td>550</td>
</tr>
<tr>
<td>ACMS2</td>
<td>220</td>
<td>220</td>
<td>550</td>
<td>550</td>
<td>550</td>
</tr>
<tr>
<td>ACMS3</td>
<td>234</td>
<td>234</td>
<td>550</td>
<td>550</td>
<td>550</td>
</tr>
<tr>
<td>ACMS4</td>
<td>256</td>
<td>256</td>
<td>550</td>
<td>550</td>
<td>550</td>
</tr>
<tr>
<td>ACMS5</td>
<td>279</td>
<td>279</td>
<td>550</td>
<td>550</td>
<td>550</td>
</tr>
<tr>
<td>ACMS6</td>
<td>294</td>
<td>294</td>
<td>550</td>
<td>550</td>
<td>550</td>
</tr>
<tr>
<td>Long-chain ACS family</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACCL1</td>
<td>275</td>
<td>275</td>
<td>550</td>
<td>550</td>
<td>550</td>
</tr>
<tr>
<td>ACCL2</td>
<td>223</td>
<td>223</td>
<td>550</td>
<td>550</td>
<td>550</td>
</tr>
<tr>
<td>ACCL3</td>
<td>296</td>
<td>296</td>
<td>550</td>
<td>550</td>
<td>550</td>
</tr>
<tr>
<td>ACCL4</td>
<td>316</td>
<td>316</td>
<td>550</td>
<td>550</td>
<td>550</td>
</tr>
<tr>
<td>ACCL5</td>
<td>300</td>
<td>300</td>
<td>550</td>
<td>550</td>
<td>550</td>
</tr>
<tr>
<td>Very long-chain ACS family</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACCLVI</td>
<td>242</td>
<td>242</td>
<td>550</td>
<td>550</td>
<td>550</td>
</tr>
<tr>
<td>ACCLVII</td>
<td>229</td>
<td>229</td>
<td>550</td>
<td>550</td>
<td>550</td>
</tr>
<tr>
<td>ACCLVIII</td>
<td>294</td>
<td>294</td>
<td>550</td>
<td>550</td>
<td>550</td>
</tr>
<tr>
<td>Bubblegum ACS family</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACSCBG1</td>
<td>281</td>
<td>281</td>
<td>550</td>
<td>550</td>
<td>550</td>
</tr>
<tr>
<td>Other ACSs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACSSF1</td>
<td>289</td>
<td>289</td>
<td>550</td>
<td>550</td>
<td>550</td>
</tr>
<tr>
<td>ACSSF2</td>
<td>262</td>
<td>262</td>
<td>550</td>
<td>550</td>
<td>550</td>
</tr>
<tr>
<td>ACSSF3</td>
<td>201</td>
<td>201</td>
<td>550</td>
<td>550</td>
<td>550</td>
</tr>
<tr>
<td>ACSSF4</td>
<td>197</td>
<td>197</td>
<td>550</td>
<td>550</td>
<td>550</td>
</tr>
<tr>
<td>Consensus</td>
<td>[Y,F]</td>
<td>[T,S]</td>
<td>[T,S]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

ACS, acyl-coenzyme A synthetase; ACSBG, bubblegum ACS; ACSF, ACS family; ACCL, long-chain ACS; ACMS, medium-chain ACS; ACSS, short-chain ACS; ACSVL, 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 fragment of the Motif III consensus found in the ACSS 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).
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).

ACS transcript variants

Multiple isoforms of seven human ACSs were detected by BLAST searching. Two variants each of human ACSS2,
TABLE 2. Human acyl-CoA synthetase family nomenclature

<table>
<thead>
<tr>
<th>Proposed Designation</th>
<th>Approved Gene Symbol</th>
<th>Approved Gene Name</th>
<th>Alternative Symbols and Aliases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short-chain ACS family</td>
<td>ACS1</td>
<td>Acyl-CoA synthetase short-chain family member 1</td>
<td>ACS2L; AceCS2L; MGC38845</td>
</tr>
<tr>
<td></td>
<td>ACS2</td>
<td>Acyl-CoA synthetase short-chain family member 2</td>
<td>ACS2; ACS; ACSA; AceCS; dJ1161H23.1</td>
</tr>
<tr>
<td></td>
<td>ACS3</td>
<td>Acyl-CoA synthetase short-chain family member 3</td>
<td>FLJ21963</td>
</tr>
<tr>
<td>Medium-chain ACS family</td>
<td>ACSM1</td>
<td>Acyl-CoA synthetase medium-chain family member 1</td>
<td>BUCS1; MACS1</td>
</tr>
<tr>
<td></td>
<td>ACSM2A</td>
<td>Acyl-CoA synthetase medium-chain family member 2A</td>
<td>LOC123876; A-923A4.1</td>
</tr>
<tr>
<td></td>
<td>ACSM2B</td>
<td>Acyl-CoA synthetase medium-chain family member 2B</td>
<td>HXMA; HYST1046</td>
</tr>
<tr>
<td></td>
<td>ACSM3</td>
<td>Acyl-CoA synthetase medium-chain family member 3</td>
<td>SAH; SA; LOC341392</td>
</tr>
<tr>
<td></td>
<td>ACSM4</td>
<td>Acyl-CoA synthetase medium-chain family member 4</td>
<td>FLJ20581</td>
</tr>
<tr>
<td></td>
<td>ACSM5</td>
<td>Acyl-CoA synthetase medium-chain family member 5</td>
<td>ACS3; ACS5; FLA5</td>
</tr>
<tr>
<td>Long-chain ACS family</td>
<td>ACSL1</td>
<td>Acyl-CoA synthetase long-chain family member 1</td>
<td>ACS1; LACS; FACL1; FACL2; LACS1; LACS2</td>
</tr>
<tr>
<td></td>
<td>ACSL2</td>
<td>Acyl-CoA synthetase long-chain family member 2</td>
<td>ACS5; FACL3; PRO2194</td>
</tr>
<tr>
<td></td>
<td>ACSL3</td>
<td>Acyl-CoA synthetase long-chain family member 3</td>
<td>ACS4; FACL4; LACS4; MRX63</td>
</tr>
<tr>
<td></td>
<td>ACSL4</td>
<td>Acyl-CoA synthetase long-chain family member 4</td>
<td>ACS2; ACS5; FLA5</td>
</tr>
<tr>
<td></td>
<td>ACSL5</td>
<td>Acyl-CoA synthetase long-chain family member 5</td>
<td>ACS2; ACS; FACL5</td>
</tr>
<tr>
<td></td>
<td>ACSL6</td>
<td>Acyl-CoA synthetase long-chain family member 6</td>
<td>ACS2; FACL6; LACS2; LACS5; KIAA0837</td>
</tr>
<tr>
<td>Very long-chain ACS family</td>
<td>ACSVL1</td>
<td>Solute carrier family 27 (fatty acid transporter), member 1</td>
<td>VLCS; FATP2; VLACS; FACVL1; hFACVL1; HsT17226</td>
</tr>
<tr>
<td></td>
<td>ACSVL2</td>
<td>Solute carrier family 27 (fatty acid transporter), member 2</td>
<td>SATP6; FACVL2; VLCS-H1</td>
</tr>
<tr>
<td></td>
<td>ACSVL3</td>
<td>Solute carrier family 27 (fatty acid transporter), member 3</td>
<td>FATP3; VLCS-3; MGC34565</td>
</tr>
<tr>
<td></td>
<td>ACSVL4</td>
<td>Solute carrier family 27 (fatty acid transporter), member 4</td>
<td>FATP; FATP1; FLJ00336; MGC71751</td>
</tr>
<tr>
<td></td>
<td>ACSVL5</td>
<td>Solute carrier family 27 (fatty acid transporter), member 5</td>
<td>FATP4</td>
</tr>
<tr>
<td></td>
<td>ACSVL6</td>
<td>Solute carrier family 27 (fatty acid transporter), member 6</td>
<td>FATP5; FACVL3; VLACSR; VLCSH2; VLCS-H2; FLJ22987</td>
</tr>
<tr>
<td>Bubblegum ACS family</td>
<td>ACSSB1</td>
<td>Acyl-CoA synthetase bubblegum family member 1</td>
<td>BG1; BGM; FLJ30329; MGC14352</td>
</tr>
<tr>
<td></td>
<td>ACSSB2</td>
<td>Acyl-CoA synthetase bubblegum family member 2</td>
<td>BGR; PRD-NY3; DKFZp443K1635</td>
</tr>
<tr>
<td>Other ACSs (ACS Family)</td>
<td>ACSS1</td>
<td>Acetoacetyl-CoA synthetase</td>
<td>SUR-5; FLJ2389</td>
</tr>
<tr>
<td></td>
<td>ACSS2</td>
<td>2-Aminoacyl-6-semialdehyde dehydrogenase</td>
<td>FLJ20920</td>
</tr>
</tbody>
</table>

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 ACSL2 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 ACSL2_v2 is shorter than that encoded by ACSL2_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 exon 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.
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 ACSS2 as well as ACSSBG 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 ± 13% (mean ± SD) for 44 alignments], whereas the identity of nonsubfamily pairs ranged from 15% to 27% (averaging 20 ± 2% for 281 alignments) (see supplementary Table I). Although the average percentage identity within subfamilies was higher for Motif II (71 ± 13% for 44 alignments) than for the full-length sequences, there was a broader range (39–97%). The average identity of nonsubfamily pairs (32 ± 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 other species.
Fig. 2. Alignment of newly identified human ACS amino acid sequences. Predicted sequences of human ACS3, ACS4, 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 199 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, ACSSL, 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 Drosophila 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 ACSS 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 >95% bootstrap support, indicating a robust estimate of the topology. The ACSF4 clade included an ancestral node with 95% bootstrap support, with a Drosophila member but no C. elegans members, and additional BLAST searching revealed no apparent medium-chain family members in other insect or nematode species.
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 ACS remains limited, particularly with respect to the mammalian enzymes. Mutagenesis experiments 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 (ACSAs; members of the ACSS 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 or domains that may be important for protein-protein interactions.

### Motif III

- ACS3: DCTVLKPTMLA
- ACS4: DCTVLKPTMLA

### Motif II

- ACS3: DCTVLKPTMLA
- ACS4: DCTVLKPTMLA

### Motif V

- ACS3: KDHNLKMIK
- ACS4: KDHNLKMIK

### Figure 2.—Continued.
**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-aminoacidipic 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 and the L-, A-, and G-motifs. The P-loop residues are 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.

Hisanaga et al. (13) defined four structurally significant domains in tACSs 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 tACSs 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, YGXTE), 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-VPMFHVNAY-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 PKTRSGXXXR (50). This motif, with consensus PKTX[S,T]GKIX[R,K], can be found in 13 human ACS sequences, including all ACSS and ACSSM 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.
unpublished observation). To the best of our knowledge, only one bona fide ACS, human ACSBG2, has a different residue (histidine) at this position. ASCBG2 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-VPMMFHVN234) 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. Maiguel 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...
acid was found on the surface of the bacterial ACSAε 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 ~140°) 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 ACS11 and ACS6 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). Soupe and Kuipers (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


