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[†]

Patricia C. Babbitt* and George L. Kenyon
Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94143

Brian M. Martin
Molecular Neurogenetics Units, Clinical Neuroscience Branch, National Institute of Mental Health, Bethesda, Maryland 20892

Hugues Charest and Michel Slyvestre
Universite-du Quêbec, INRS-Santê, 245 boul. Hymus Pointe-Claire, Quêbec, H9R 1G6, Canada

Jeffrey D. Scholten,[‡] Kai-Hsuan Chang, Po-Huang Liang, and Debra Dunaway-Mariano*
Department of Chemistry and Biochemistry, University of Maryland, College Park, Maryland 20742

Received January 30, 1992; Revised Manuscript Received March 27, 1992

ABSTRACT: We have deduced the nucleotide sequence of the genes encoding the three components of 4-chlorobenzoate (4-CBA) dehalogenase from *Pseudomonas* sp. CBS-3 and examined the origin of these proteins by homology analysis. Open reading frame 1 (ORF1) encodes a 30-kDa 4-CBA-coenzyme A dehalogenase related to enoyl-coenzyme A hydratases functioning in fatty acid β-oxidation. ORF2 encodes a 57-kDa protein which activates 4-CBA by acyl adenylation/thioesterification. This 4-CBA:coenzyme A ligase shares significant sequence similarity with a large group of proteins, many of which catalyze similar chemistry in β-oxidation pathways or in siderophore and antibiotic synthetic pathways. These proteins have in common a short stretch of sequence, (T,S)(S,G)G(T,S)(T,E)G(L,X)PK(G,–), which is particularly highly conserved and which may represent an important new class of "signature" sequence. We were unable to find any proteins homologous in sequence to the 16-kDa 4-hydroxybenzoate-coenzyme A thioesterase encoded by ORF3. Analysis of the chemistry and function of the proteins found to be structurally related to the 4-CBA:coenzyme A ligase and the 4-CBA-coenzyme A dehalogenase supports the proposal that they evolved from a β-oxidation pathway.

Because of their widespread use as industrial and agricultural agents, halogenated hydrocarbons constitute a particularly formidable class of environmental pollutants. Conventional chemical methods of disposal/detoxification are both costly and inefficient, prompting intensive efforts to find better alternatives. One such alternative, biodegradation of these compounds by microorganisms, offers a promising approach to the detoxification of contaminated areas [for recent reviews see Abramowicz (1990) and Commandeur and Parsons (1990)]. In recent years a number of strains of soil-dwelling bacteria have been isolated which are able to catabolize a variety of halogenated hydrocarbons. One of these strains, *Pseudomonas* sp. CBS-3, was isolated by requiring growth on 4-chlorobenzoate (4-CBA)[1] as the sole source of carbon (Keil et al., 1981). The 4-CBA is metabolized in this bacterium first by conversion to 4-hydroxybenzoate (4-HBA) and then to 3,4-dihydroxybenzoate. This latter metabolite is further degraded via the ortho-cleavage and β-ketoadipate pathways (Ornston, 1990).

The novel enzyme system of the 4-CBA biodegradative pathway is the 4-CBA dehalogenase which catalyzes an unprecedented aromatic substitution reaction involving the replacement of the chloride substituent with the hydroxyl group from a molecule of water (Müller et al., 1984). Interestingly, in addition to the 4-CBA dehalogenase, the *Pseudomonas* sp. CBS-3 strain has also been shown to contain (i) a two-component dioxygenase (Klages et al., 1981; Markus et al., 1984) which converts 4-chlorophenylacetate to the metabolite 3,4-dihydroxyphenylacetate and (ii) two 2-haloalkanoic acid dehalogenases (Klages et al., 1983; Schneider et al., 1991) which catalyze the hydrolytic dehalogenation of 2-monochloroacetate and 2-monochloropropionate. The occurrence of three unique dehalogenation pathways in a single bacterial strain (screened solely for the ability to convert 4-CBA to 4-HBA) is curious indeed.

Previous work led to the cloning and expression of the *Pseudomonas* sp. CBS-3 4-CBA dehalogenase genes in *Escherichia coli* (Savard et al., 1986) and to the identification of the gene translation products (Scholten et al., 1991). The 4-CBA dehalogenase was found to consist of a 4-CBA:CoA ligase, a 4-CBA-CoA dehalogenase, and a 4-HBA-CoA thioesterase (Scholten et al., 1991; Chang et al., 1992). The

[†] This work was supported by NIH Grant GM-28688 to D.D.-M. by NIH Grants AR17323 and GM40570 to G.L.K., and by the National Sciences and Engineering Research Council of Canada by the fonds pour la Formation de Chercheurs l'Aide ã la Reserche du Quebec to M.S.

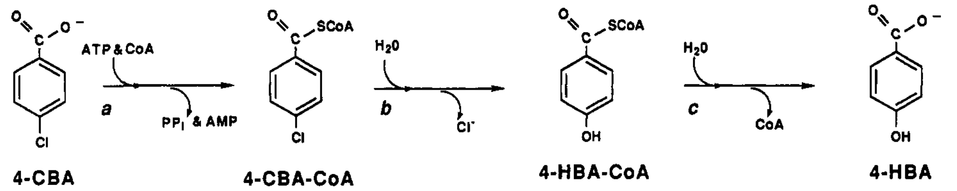
* To whom correspondence should be addressed.

[‡] Present address: Parke-Davis, Pharmaceutical Research Division, Warner-Lambert Company, Ann Arbor, MI 48106-1047.

[1] Abbreviations: 4-CBA, 4-chlorobenzoate; 4-HBA, 4-hydroxybenzoate; CoA, coenzyme A; SDS–PAGE, sodium dodecyl sulfate–polyacrylamide gel electrophoresis; Tris, tris(hydroxymethyl)aminomethane; Caps, 3-(cyclohexylamino)propanesulfonic acid; EMBL, European Molecular Biology Laboratory; PIR, Protein Identification Resource; ATP, adenosine 5′-triphosphate; AMP, adenosine 5′-monophosphate; NADPH, dihydronicotinamide adenine dinucleotide phosphate; ORF, open reading frame.

0006-2960/92/0431-5594$03.00/0

Scheme I: Reaction Steps of the Dehalogenation of 4-CBA in *Pseudomonas* sp. CBS-3 Catalyzed by (a) 4-CBA:CoA Ligase, (b) 4-CBA-CoA Dehalogenase, (c) 4-HBA-CoA Thioesterase



goal of the present work was to determine the amino acid sequences of the proteins comprising the 4-CBA dehalogenase of *Pseudomonas* sp. CBS-3 and to examine their origin by homology analysis. In this paper, the amino acid sequences of the dehalogenase polypeptide components are reported as determined by gene subcloning and sequencing. The relationships found to exist between the 4-CBA:CoA ligase and families of acyl:adenyl ligases and between the 4-CBA-CoA dehalogenase and the enoyl-CoA hydratases/isomerases of fatty acid β-oxidation pathways are described. In addition, an analysis of the primary structure of the 4-HBA-CoA thioesterase in relation to other thioesterases and acyl carrier proteins is performed.

Materials and Methods

Materials. The Sequenase kit and buffers were obtained from U.S. Biochemical Corp. The nested deletions kits were purchased from Amersham and Pharmacia as were the M13 sequencing kit and [α-^{35}S]dATPαS. Restriction enzymes and T4 DNA ligase were obtained from Bethesda Research Laboratories and Promega. All other reagents used were obtained from Aldrich or Sigma Chemical Companies.

Sequencing Strategy. Sequencing was carried out using a 4.5-kb chromosomal DNA fragment cloned from *Pseudomonas* sp. CBS-3 into *E. coli* on plasmid pMMB22 (Scholten et al., 1991). The 4.5-kb insert was cut from the plasmid and into two segments (1.6 kb and 3.0 kb) by *Sma*I and *Sal*I digestion and sequenced as illustrated in Figure 1 of the supplementary material. The 3.0-kb segment was subcloned into pUC19 and analyzed by using a sequenase kit in conjunction with synthetic and universal primers. Oligonucleotide primers were synthesized with a Biosearch DNA Synthesizer (Model 8750). Subclones used for sequencing with the M13 universal forward and reverse primers were generated by using the method of nested deletions (Henikof, 1984). Sequencing was accomplished using the dideoxy chain termination method with the modified form of T7 DNA polymerase (Tabor & Richardson, 1987) and [α-^{35}S]dATPαS. The 1.6-kb fragment was sequenced by constructing *Eco*RI, *Pst*I, *Sph*I, and *Hin*dIII deletion clones in M13mp18 or M13mp19. Both universal and synthetic primers were used in conjunction with an Amersham M13 sequencing kit (shown in Figure 1 of the supplementary material).

N-Terminal Polypeptide Sequencing. The three polypeptides were isolated from *E. coli* subclones as described in the following paper (Chang et al., 1992) and chromatographed on 12% acrylamide SDS–PAGE gels using a solution containing 3 g/L Tris base, 14.4 g/L glycine, and 1 g/L SDS in water as the running buffer. Dried gels were incubated in 10 mM K$^+$Caps buffer (pH 11), containing 10% methanol for 10 min, and then transfer-blotted on poly(vinylidene difluoride) membranes with a Hoefer Semiphor transfer blotter (0.8 nA/cm^2; 45 min). After blotting, the membrane was stained with Coomassie blue and the protein bands were cut from the gel and subjected to Edman degradation in an Applied Biosystems 470A gas-phase protein sequenator. The sequences determined for the polypeptide N-terminals are as follows: 4-CBA:CoA ligase (57-kDa polypeptide), Met-Gln-Thr-Val-Glu-Met; 4-CBA-CoA dehalogenase (30-kDa polypeptide), Met-Tyr-Glu-Ala-Ile-Gly; 4-HBA-CoA thioesterase (16-kDa polypeptide), Met-Ala-Arg-Ser-Ile-Thr-Met-Gln-Gln-Arg-Ile-Glu-Phe-Gly-Asp.

Primary Sequence Homology Analysis. The three open reading frames of the dehalogenase were compared to proteins in the Protein Identification Resource (PIR), Release No. 28 (George et al., 1986) and the database compiled by the European Molecular Biology Laboratory (EMBL) Release No. 18 (Hamm & Cameron., 1986) using the FASTA program (Pearson & Lipman, 1988) and the programs available in the EuGene Interface of the Molecular Biology Information Resource (Shalom et al., 1989). Statistical analysis was performed using algorithms from EuGene (Altschul & Erickson, 1986; Lawrence & Goldman 1988). Alignments were generated using the PIMA algorithm (Smith & Smith, 1990) and then optimized by inspection. On the basis of the results of the database search and inferences from chemical intuition, the literature was then examined for other likely relatives not included in the databases. These candidates were screened for statistically significant relationships to the sets of known homologous proteins using the program described.

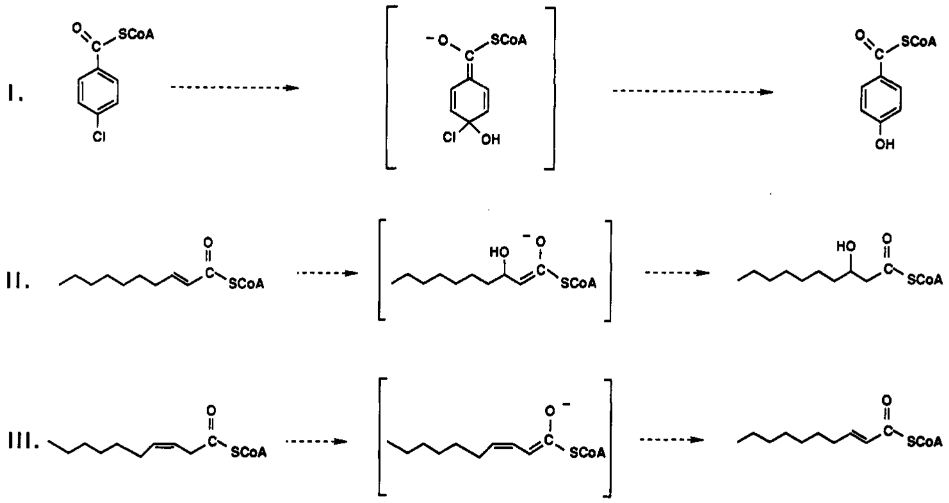
Results and Discussion

Previous studies of the *Pseudomonas* sp. CBS-3 4-CBA dehalogenase system in one of our laboratories resulted in the cloning of a 9.8-kb fragment of chromosomal DNA in pMMB22 (Savard et al., 1986) which conferred on *E. coli* the ability to convert 4-CBA to 4-HBA. Circumscription of the genes to a 4.5-kb DNA fragment and analysis of the gene translation products revealed three polypeptides which were sized at 57, 30, and 16 kDa, respectively (Scholten et al., 1991). The 57-kDa polypeptide (4-CBA:CoA ligase) was found to catalyze, as illustrated in Scheme I, the adenylation and then thioesterification of 4-CBA with ATP and CoA, respectively. The 30-kDa polypeptide (4-CBA-CoA dehalogenase) was discovered to catalyze the hydrolytic dechlorination of 4-CBA-CoA to 4-HBA-CoA (Scheme I). Finally, the 16-kDa polypeptide (4-HBA-CoA thioesterase) was shown to catalyze the hydrolysis of 4-HBA-CoA to generate 4-HBA and CoA (Scheme I).

Characterization of the 4-Chlorobenzoate Dehalogenase Genes. Oligonucleotide sequencing of the 4.5-kb DNA fragment was undertaken to locate the dehalogenase genes and to determine the amino acid sequences of the gene products. The 4.5-kb DNA fragment was sequenced as twc separate pieces. The sequencing strategy used is shown in Figure 1 of the supplementary material while the nucleotide sequence of the encoding region of the 4.5-kb fragment thus obtained is shown in Figure 1 of the text. The G/C ratio for this DNA fragment was determined to be 58%.

Three opening reading frames (ORFs) corresponding to 29 847, 57 155, and 16 107 Da polypeptides arranged contiguously 5′ to 3′ on the 4.5-kb fragment were identified. The noncoding intervals between the ~30-kDa and the ~57 kDa ORFs and between the ~57 kDa and 16 kDa ORFs comprise

Scheme II: Parallel Reaction Pathways Catalyzed by 4-CBA-CoA Dehalogenase (I), 2-Enoyl-CoA Hydratase (II), and Δ^3-*cis*-Δ^2-*trans*-Enoyl-CoA Isomerase (III)



8 and 105 bases, respectively. The ORFs were verified by comparing the predicted size and N-terminal sequences of the encoded polypeptides (Figure 1) with the molecular weights and N-terminal amino acid sequences (see Materials and Methods) determined by SDS–PAGE analysis (Chang et al., 1992) and Edman degradation of the three dehalogenase polypeptide components purified from the *E. coli* clone. The purine-rich regions corresponding to possible ribosome binding sites are shown in Figure 1. We were unable to locate an obvious promoter site.

Primary Structure Relationships between the 30-kDa Polypeptide, 4-CBA-CoA Dehalogenase, and Other Proteins. A comparison of the amino acid sequence of the 30-kDa 4-CBA-CoA dehalogenase encoded by ORF1 (Figure 1) with the protein sequences contained in the PIR and EMBL databases was made to probe its origin. Four structurally related enzymes (Palosaari et al., 1991; Minami-Ishii et al., 1989), each of which functions in the fatty acid β-oxidation pathway, were identified as sharing significant sequence identity with the 4-CBA-CoA dehalogenase. Two are monofunctional enzymes found in rat liver mitochondria, the 2-enoyl-CoA hydratase (Minami-Ishii et al., 1989), and the Δ^3-*cis*-Δ^2-*trans*-enoyl-CoA isomerase (Palosaari et al., 1991). The other two are the 2-enoyl-CoA hydratase domains of the trifunctional enzyme of rat liver peroxisomes (Osumi et al., 1985) and the *E. coli* multifunctional enzyme (Dirusso et al., 1990; Nakahigashi & Inokuchi, 1990; Yang et al., 1991). The hydratase domain of the trifunctional rat liver enzyme also exhibits Δ^3-*cis*-Δ^2-*trans*-enoyl-CoA isomerase activity (Ishii et al., 1987; Minami-Ishii et al., 1989; Palosaari & Hiltunen., 1990; Palosaari et al., 1991) while the hydratase domain of the bacterial multifunctional enzyme exhibits 3-hydroxyacyl-CoA epimerase activity (Yang et al., 1991). The other domains of the two multifunctional enzymes, which catalyze the oxidation and carbon–carbon bond cleavage steps of the fatty acid degradative pathway, are not related in sequence to either the hydratase domain or the dehalogenase.

A possible link between the mechanisms of action of the 2-enoyl-CoA hydratase of the fatty acid β-oxidation pathway and the dehalogenase of the 4-CBA dechlorination pathway is apparent from the similarities in the reactions that they catalyze (see Scheme II). Both the hydratase and dehalogenase activate water for addition across a carbon–carbon bond that is in conjunction with the CoA-thioester group. The reactions differ in that H_2O addition to the hydratase substrate occurs at the β-carbon in a Michael process (reaction II, Scheme II) while in the dehalogenase substrate it occurs at the para carbon of the aromatic ring in a formal 1,6-addition process (reaction I, Scheme II). For the latter reaction, the presence of a good leaving group (Cl^-) coupled with thermodynamically driven aromatization results in a spontaneous re-formation of the carbon–carbon double bond and retention of the hydroxyl substituent in the 4-HBA-CoA product. Hence, while the hydratase catalyzes the addition of H_2O across the carbon–carbon double bond, the dehalogenase catalyzes H_2O addition and Cl^- elimination. The chemical focus of these two enzymes would, however, seem to be the same, i.e., to activate the water molecule for nucleophilic addition and, by protonation or ion pair formation, to stabilize the enolate formed upon addition.[2]

Catalysis by the Δ^3-*cis*-Δ^2-*trans*-enoyl-CoA isomerase involves positional and geometric isomerization of the 3-cis to the 2-trans bond of a 3-*cis*-enoyl CoA ester (reaction III, Scheme II). Although the mechanisms of the Δ^3-*cis*-Δ^2-*trans*-enoyl-CoA isomerase reaction has not been well characterized (Euler-Bertram & Stoffel., 1990), in analogy to the well-defined mechanism of 3-oxo-Δ^5-steroid isomerase (Hawkinson et al., 1991; Kuliopulos et al., 1991), one could envision allylic deprotonation at C(2) to generate the dienol(ate) intermediate, followed by protonation of this intermediate at C(4) to generate the Δ^2-*trans*-enoyl-CoA product (reaction III, Scheme II). Catalysis by the Δ^3-*cis*-Δ^2-*trans*-enoyl-CoA isomerase may thus entail stabilization of a CoA thioester enolate in a catalytic strategy similar to that employed by the enoyl-CoA hydratase and 4-CBA-CoA dehalogenase.

An alignment optimizing identity between the amino acid sequences of the 4-CBA-CoA dehalogenase, the isomerase, and the three hydratases is shown in Figure 2. Statistical analysis suggests that the similarities are significant, falling in the range indicative of either "possible relationship" or "probable relationship" between each pair of sequences. The overall sequence identities beween protein pairs fall between

[2] Kinetic isotope studies carried out with the monofunctional 2-enoyl-CoA hydratase (crotonase) from bovine liver (Bahnson & Anderson, 1991) suggest that protonation at C(2) occurs in concert with the addition of water at C(3). In this case, stabilization of an enolate intermediate would not be necessary. On the other hand, formation of an enol or enolate intermediate during turnover in the 4-CBA-CoA dehalogenase active site is unavoidable. Thus, either the catalytic mechanisms of the hydratase and dehalogenase have diverged along with their amino acid sequences or the hydratase does, in fact, catalyze the hydration of its substrate, crotonate, by a stepwise rather than a concerted mechanism.

17% and 30% with some segments of the sequences exhibiting much higher sequence similarity than others, particularly in the regions shaded black or grey in Figure 2. These regions, along with the several residues conserved throughout all five sequences (indicated by an asterisk in Figure 2), may preserve the scaffold of a tertiary structure and/or a catalytic apparatus common to each of these enzymes. On the basis of this alignment and the chemical evidence, we suggest that the 30-kDa protein of the dehalogenase system shares a common ancestry with the enoyl-CoA hydratases of the fatty acid β-oxidation pathways of mitochondrial, eukaryotic peroxisomal, and bacterial origins.

In parallel with these findings, we expected to find relationships between this set of proteins and other proteins performing similar functions in other species. Contrary to these expectations, however, we could find no such relationship with the peroxisomal trifunctional (hydratase–dehydrogenase–epimerase) enzyme functioning in fatty acid β-oxidation in the yeast *Candida tropicalis* (Nuttley et al., 1988). We were also unable to find significant similarities between the protein sequences shown in Figure 2 and the PIR/EMBL collection of sequences of proteins which catalyze cleavage reactions in β-keto-CoA thioesters. Attempted alignments of the HMG-CoA lyase, acetyl-CoA acetyl transferase, 3-ketoacyl-CoA thiolase, and citrate synthetase sequences failed. Proteins catalyzing reactions of carboxylate substrates proceeding through aci-acid intermediates were also found to be unrelated. This latter group includes mandelate racemase, fumerase, aconitase, and enolase. Nor were the lyases aspartate ammonia lyase, argininosuccinate lyase, or 3-hydroxyl-3-methylglutaryl lyase found to be structurally related. Finally, the 4-CBA-CoA dehalogenase does not appear to be related in sequence to the two 2-haloalkanoic dehalogenases of the *Pseudomonas* sp. CBS-3 (Schneider et al., 1991). Thus, on the basis of available sequence information, the 4-CBA-CoA dehalogenase appears to be related to the 2-enoyl-CoA hydratase and the Δ^3-*cis*-Δ^2-*trans*-enoyl-CoA isomerase of one family of fatty acid β-oxidation pathway enzymes and to no other specific group of enzymes which are functionally or mechanistically similar.[3,4]

Primary Structure Relationships between the 57-kDa Polypeptide, 4-CBA:CoA Ligase, and Other Proteins. A comparison of the amino acid sequence of the 57-kDa 4-CBA:CoA ligase encoded by ORF2 with the protein sequences in the PIR and EMBL databases was made to probe the origin of this enzyme. Our previous search, carried out at the beginning of 1991, turned up six proteins as having significant sequence homology with the 4-CBA:CoA ligase (Scholten et al., 1991). As shown in Table I, this list has now grown to 25 proteins, 12 of which catalyze different reactions. This is the first report that acetate:CoA ligase and AngR, in addition to the 4-CBA:CoA ligase, are related to any of the proteins listed in Table I. The remaining sequences have been recently reported to be related to at least one and in some cases to several of the proteins included in Table I (Hori et al., 1989, 1991; Kraetzschmar et al., 1989; Diez et al., 1990; Gutiérrez et al., 1991; MacCabe et al., 1991; Rusnak et al., 1991; Morris & Jinks-Robertson, 1991; Wood et al., 1989; Becker-Andrê et al., 1991; Suzuki et al., 1990; Toh et al., 1990). In the present study, a subset of 13 of these proteins, representing a wide range of different reactions, substrates, and metabolic functions, was aligned with the 4-CBA:CoA ligase. The region of the alignment showing the highest degree of similarity among these sequences (spanning approximately the C-terminal half of the 4-CBA-CoA ligase sequence) is shown in Figure 3.

Table I: Protein Sequences Homologous to 4-Chlorobenzoate:CoA Ligase[a]

protein	species	reference
gramicidin S synthetase 1	*Bacillus brevis* (Nagano)	Hori et al. (1989)
	Bacillus brevis (ATCC 9999)	Kraetzschmar et al. (1989)
gramicidin S synthetase 2	*Bacillus brevis*	Hori et al. (1991)
tyrocidin synthetase 1	*Bacillus brevis*	Weckermann et al. (1988)
α-aminoadipyl–cysteinyl–valine (ACV) synthetase[b]	*Penicillium chryosogenum*	Diez et al. (1990)
	Cephalosporium acremonium	Gutierrez et al. (1991)
	Aspergillus nidulans	MacCabe et al. (1991)
enterobactin synthetase component F	*Escherichia coli*	Rusnak et al. (1991)
angR gene product	*Vibrio anguillarum*	Farrell et al. (1990)
α-aminoadipate reductase	*Saccharomyces cerevisiae*	Morris et al. (1991)
photinus–luciferin 4-monooxygenase	*Photinus pyralis*	de Wet et al. (1987)
	Pyrophorus plagiophthalamus[c]	Wood et al. (1989)
4-coumarate:CoA ligase[d]	*Petroselinum crispum*	Lozoya et al. (1988)
	Oryza sativa	Zhao et al. (1990)
	Solanum tuberosum	Becker-Andrê et al. (1991)
long-chain fatty acid:CoA ligase	*Rattus norvegicus*	Suzuki et al. (1990)
enterobactin synthetase component E	*Escherichia coli*	Staab et al. (1989)
acetate:CoA ligase	*Aspergillus nidulans*	Connerton et al. (1990)
	Neurospora crassa	Connerton et al. (1990)

[a] Sequences and references were taken from the PIR or GenBank whenever possible. When sequences were not available in the databases, the earliest publication of the sequence is cited. [b] The ACV synthetase proteins from the species listed are all composed of three homologous domains designated A, B, and C, reading from N-terminus to C-terminus. [c] There are four nearly identical sequences known for this species which are distinguished by the colors of bioluminescence they catalyze. [d] The three species expressing 4-coumarate:CoA ligase each exhibit two isoenzymes of nearly identical sequences.

[3] Analysis of the alignment shown in Figure 2 allows us to modify some conclusions on the basis of earlier alignments involving fewer sequences. Contrary to earlier claims (Minami-Ishii et al., 1989), there is no statistically significant relationship between the hydratases shown in Figure 2 and other "hydratase" family enzymes such as pig heart fumarase. A significant degree of homology has been noted, on the other hand, between the fumarases and such other hydratases as aspartate ammonia lyase (Sacchettini et al., 1988). The slight similarity to the adenine recognition loop in the CoA binding of citrate synthase (VVPGYGH) noted in an earlier alignment of the isomerase and the hydratases (AVNGYAL; residues 108–114 of the "coli.fad B" sequence in Figure 2) (Palosaari et al., 1991) shows less conservation of this former motif when the 4-CBA-CoA dehalogenase sequence is added to the alignment.

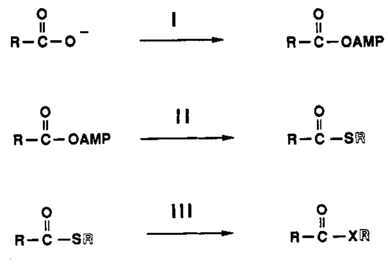
[4] Our general search of the database for sequences related to the 4-CBA-CoA dehalogenase did reveal a low degree of sequence similarity between the dehalogenase and the yeast lysyl-tRNA synthetase protein (Mirande & Waller, 1988). Comparison of this sequence to the alignment shown in Figure 2 reveals that the most highly conserved regions in the alignment (black-shaded in Figure 2) coincide with the regions of similarity with the synthetase (data not shown), leading to the conclusion that this protein may also be distantly related. Statistical analysis failed to detect a significant relationship between the synthetase and any of the proteins shown in Figure 2, however. Nor, on the basis of mechanistic or functional insights, can we rationalize a primary structure relationship between the lysyl-tRNA synthetase and the proteins shown in Figure 2.

```
              641                        661                        681                       701
CGG TGG AAT ATG CTT TAC GTC ACG GTT AGA CAG GAA TCA ACC ACG GAG GAA GAC TCA ATG TAT GAG GCA ATT GGT CAC CGC GTC GAA GAT
                                                                    Orf 1   met tyr glu ala ile gly his arg val glu asp

 721                       741                       761                         781                       801
GGT GTG GCG GAA ATT ACC ATA AAG CTT CCG CGC CAC CGG AAC GCA TTG TCG GTG AAA GCG ATG CAG GAA GTT ACG GAT GCG CTC AAT CGC
gly val ala glu ile thr ile lys leu pro arg his arg asn ala leu ser val lys ala met gln glu val thr asp ala leu asn arg

              821                        841                        861                       881
GCG GAG GAA GAC GAT TCG GTT GGC GCA GTC ATG ATC ACC GGT GCC GAG GAT GCC TTC TGT GCG GGT TTC TAT CTG CGG GAA ATC CCG CTG
ala glu glu asp asp ser val gly ala val met ile thr gly ala glu asp ala phe cys ala gly phe tyr leu arg glu ile pro leu

 901                       921                       941                         961                       981
GAC AAA GGG GTC GCC GGT GTC CGT GAC CAT TTC AGG ATC GGC GCA CTG TGG TGG CAC CAG ATG ATC CAC AAA ATT ATC CGT GTG AAG CGG
asp lys gly val ala gly val arg asp his phe arg ile gly ala leu trp trp his gln met ile his lys ile ile arg val lys arg

              1001                       1021                       1041                      1061
CCG GTA CTT GCC GCT ATC AAC GGC GTG GCG GCT GGT GGT GGA CTT GGG ATT TCG CTC GCG AGT GAC ATG GCG ATC TGT GCA GAC AGC GCA
pro val leu ala ala ile asn gly val ala ala gly gly gly leu gly ile ser leu ala ser asp met ala ile cys ala asp ser ala

 1081                      1101                      1121                        1141                      1161
AAG TTC GTC TGT GCA TGG CAC ACG ATC GGT ATC GGC AAC GAC ACA GCT ACC AGC TAC AGT CTG GCG CGT ATC GTC GGT ATG CGA CGG GCG
lys phe val cys ala trp his thr ile gly ile gly asn asp thr ala thr ser tyr ser leu ala arg ile val gly met arg arg ala

           1181                       1201                          1221                      1241
ATG GAG CTG ATG CTT ACG AAC CGG ACG CTT TAC CCG GAG GAA GCG AAG GAC TGG GGG CTC GTC AGC CGC GTA TAC CCG AAA GAT GAG TTC
met glu leu met leu thr asn arg thr leu tyr pro glu glu ala lys asp trp gly leu val ser arg val tyr pro lys asp glu phe

 1261                      1281                      1301                        1321                      1341
CGC GAA GTG GCA TGG AAA GTC GCC CGC GAA CTT GCA GCC GCT CCG ACC CAT CTC CAG GTG ATG GCG AAG GAA CGC TTC CAC GCC GGA TGG
arg glu val ala trp lys val ala arg glu leu ala ala ala pro thr his leu gln val met ala lys glu arg phe his ala gly trp

              1361                       1381                       1401                      1421
ATG CAA CCG GTC GAG GAG TGC ACC GAA TTC GAA ATT CAG AAT GTC ATC GCT TCG GTA ACG CAT CCT CAC TTC ATG CCC TGT CTT ACC AGA
met gln pro val glu glu cys thr glu phe glu ile gln asn val ile ala ser val thr his pro his phe met pro cys leu thr arg

 1441                      1461                      1481                        1501
TTC CTG GAC GGC CAT CGC GCG GAT AGG CCG CAG GTC GAA TTG CCG GCG GGC GTG TAG GAG TCC TT
phe leu asp gly his arg ala asp arg pro gln val glu leu pro ala gly val  *      Orf 2


                     1521                       1541                      1561                       1581
ATG CAG ACC GTC CAC GAG ATG CTT CGT CGG GCG GTG TCG CGT GTG CCG CAT CGC TGG GCT ATC GTC GAC GCC GCA CGC TCG ACG TTT GAC
met gln thr val his glu met leu arg arg ala val ser arg val pro his arg trp ala ile val asp ala ala arg ser thr phe asp


        1601                      1621                       1641                       1661                      1681
ATA TGT AGA ACT GGC GAG ACA AGT AGA AAC GAG GGC TCA GCA ACT GCT CGC CTG TGG CCT CAA CCC GCG CGA CCG CTT GCC GTG GTT TCG
ile cys arg thr gly glu thr ser arg asn glu gly ser ala thr ala arg leu trp pro gln pro ala arg pro leu ala val val ser

                     1701                       1721                      1741                       1761
GGC AAT TCG GTT GAG GCG GTG ATA GCC GTT CTT GCT CTT CAT CGC CTG CAG GCA GTG CCC GCG TTA ATG AAC CCA CGG CTC AAG CCG GCG
gly asn ser val glu ala val ile ala val leu ala leu his arg leu gln ala val pro ala leu met asn pro arg leu lys pro ala

        1781                      1801                       1821                       1841                      1861
GAA ATC AGT GAA CTG GTA GCA CGT GGC GAA ATG GCG CGG GCG GTG GTG GCC AAC GAT GCG GGC GTG ATG GAG GCT ATC CGG ACA CGG GTG
glu ile ser glu leu val ala arg gly glu met ala arg ala val val ala asn asp ala gly val met glu ala ile arg thr arg val

                     1881                       1901                      1921                       1941
CCG TCC GTA TGC GTT CTG GCA CTG GAC GAT CTC GTT AGC GGT TCC CGC GTC CCG GAA GTT GCC GGG AAG TCC CTC CCA CCG CCG CCG TGC
pro ser val cys val leu ala leu asp asp leu val ser gly ser arg val pro glu val ala gly lys ser leu pro pro pro pro cys

        1961                      1981                       2001                       2021                      2041
GAG CCG GAG CAG GCG GGA TTC GTT TTC TAC ACG TCG GGG ACA ACC GGT TTG CCC AAG GGA GCG GTG ATC CCC CAA CGC GCC GCC GAG AGC
glu pro glu gln ala gly phe val phe tyr thr ser gly thr thr gly leu pro lys gly ala val ile pro gln arg ala ala glu ser

                     2061                       2081                      2101                       2121
CGT GTT TTG TTT ATG GCC ACG CAG GCG GGG TTG CGG CAC GGA TCG CAT AAC GTG GTG CTC GGG TTA ATG CCT CTG TAT CAC ACA ATC GGT
arg val leu phe met ala thr gln ala gly leu arg his gly ser his asn val val leu gly leu met pro leu tyr his thr ile gly

        2141                      2161                       2181                       2201                      2221
TTC TTT GCG GTG CTG GTA GCG GCA ATG GCG TTC GAC GGG ACT TAC GTG GTT GTT GAG GAG TTC GAC GCC GGG AAC GTC CTT AAA CTA ATC
phe phe ala val leu val ala ala met ala phe asp gly thr tyr val val val glu glu phe asp ala gly asn val leu lys leu ile

                     2241                       2261                      2281                       2301
GAG CGG GAA CGC GTT ACG GCG ATG TTT GCC ACG CCG ACA CAT CTT GAC GCA CTG ACG ACA GCG GTC GAG CAG GCC GGT GCG CGG CTG GAA
glu arg glu arg val thr ala met phe ala thr pro thr his leu asp ala leu thr thr ala val glu gln ala gly ala arg leu glu

        2321                      2341                       2361                       2381                      2401
TCG CTA GAG CAC GTG ACT TTC GCG GGC GCC ACG ATG CCG GAC ACG GTG CTC GAA AGA GTC AAT CGT TTT ATT CCG GGA GAG AAA GTC AAC
ser leu glu his val thr phe ala gly ala thr met pro asp thr val leu glu arg val asn arg phe ile pro gly glu lys val asn
```

```
                     2421                       2441                      2461                      2481
ATC TAC GGA ACA ACA GAA GCG ATG AAT TCG CTG TAC ATG CGC GCC GTC CGC ATA GCC GGC ACT GTG ATG CGT CCT GGC TTT TTT TCT GAA
ile tyr gly thr thr glu ala met asn ser leu tyr met arg ala val arg ile ala gly thr val met arg pro gly phe phe ser glu

        2501                      2521                       2541                      2561                      2581
GTT CGA ATA GTG CGC GTT GGC GGC GAC GTT GAC GAC GGT TGC CCG ACG GTG AAG AGG GCG AGC TGG CGG TGG CGG CGA CGG ATG CGA CCT
val arg ile val arg val gly gly asp val asp asp gly cys pro thr val lys arg ala ser trp arg trp arg arg arg met arg pro

                     2601                       2621                      2641                      2661            TTT
CAG GCT ACC TTA ACC AAC CTG AGG CTA CTG CAG AAA AGC TTC AGA AAG GCT GGT ACC GGA CGA GCG ATA TGC GTG CGG GAC GGC AGT
phe gln ala thr leu thr asn leu arg leu leu gln lys ser phe arg lys ala gly thr gly arg ala ile cys val arg asp gly ser

        2681                      2701                       2721                      2741                      2761
GGC AAC ATC GTC GTC CTC GGC CGC GTG GAC GAC ATG ATC ATT TCC GGT GGT GAA AAC ATC CAT CCG TCG GAA GTG GAA CGG ATT CTG GCG
gly asn ile val val leu gly arg val asp asp met ile ile ser gly gly glu asn ile his pro ser glu val glu arg ile leu ala

                     2781                       2801                      2821                      2841
GCG GCG CCG GGC GTC GCT GAA GTC GTC GTG ATA GGT GTG AAA GAC GAA CGG TGG GGA CAA AGC GTT GTT GCC TGC GTG GTG TTG CAG CCG
ala ala pro gly val ala glu val val val ile gly val lys asp glu arg trp gly gln ser val val ala cys val val leu gln pro

        2861                      2881                       2901                      2921                      2941
GGT GCG TCC GCG TCG GCT GAA CGG CTG GAT GCC TTT TGT CGC GCG AGC GCA CTT GCC GAC TTC AAG CGA CCG CGC CGA TAC GTG TTC CTG
gly ala ser ala ser ala glu arg leu asp ala phe cys arg ala ser ala leu ala asp phe lys arg pro arg arg tyr val phe leu

                     2961                       2981                      3001                      3021
GAC GAG TTG CCG AAA AGC GCC ATG AAC AAA GTC CTG CGC AGA CAA CTC ATG CAG CAC GTG AGC GCG ACT TCC AGT GCC GCG GTA GTG CCG
asp glu leu pro lys ser ala met asn lys val leu arg arg gln leu met gln his val ser ala thr ser ser ala ala val val pro

        3041                      3061                       3081                      3101                      3121
GCG CCA GCG GTA AAG CAG AGA ACA TAT GCT CCC TCG GGA CGC GCC ATC GCC CGC TAG CTG GAC AAG CTA GAA CTA CGG TGT GCG ACG CAC
ala pro ala val lys gln arg thr tyr ala pro ser gly arg ala ile ala arg  *

                     3141                       3161                      3181
TCC GGA GCG AAT TGC ATG GGC AGC GTA GCA GGG GAT TCC GAA TAC AAG TGG ATG AAA AAT


                     3201                       3221                      3241                      3261
GGG ACT GCG GTC ATG GCA CGA TCA ATA ACT ATG CAG CAA CGG ATC GAA TTC GGC GAC TGT GAT CCG GCG GGC ATC GTG TGG TTT CCG AAT
        Orf 3   met ala arg ser ile thr met gln gln arg ile glu phe gly asp cys asp pro ala gly ile val trp phe pro asn

        3281                      3301                       3321                      3341                      3361
TAC CAC CGC TGG CTG GAC GCT GCG TCG AGA AAC TAT TTC ATC AAG TGC GGT CTA CCT CCA TGG AGA CAG ACC GTT GTG GAA CGA GGG ATC
tyr his arg trp leu asp ala ala ser arg asn tyr phe ile lys cys gly leu pro pro trp arg gln thr val val glu arg gly ile

                     3381                       3401                      3421                      3441
GTT GGT ACC CCG ATC GTC AGT TGC AAT GCA TCG TTC GTG TGC ACG GCG TCC TAC GAC GAC GTT TTA ACC ATC GAA ACA TGC ATC AAG GAA
val gly thr pro ile val ser cys asn ala ser phe val cys thr ala ser tyr asp asp val leu thr ile glu thr cys ile lys glu

        3461                      3481                       3501                      3521                      3541
TGG CGG CGC AAA AGC TTC GTC CAA CGG CAC AGC GTT TCT AGG ACG ACT CCG GGC GGG GAC GTA CAG CTA GTA ATG CGT GCG GAC GAA ATC
trp arg arg lys ser phe val gln arg his ser val ser arg thr thr pro gly gly asp val gln leu val met arg ala asp glu ile

                     3561                       3581                      3601                      3621
CGG GTT TTT GCG ATG AAC GAC GGC GAA CGG CTA CGT GCG ATT GAA GTC CCC GCG GAC TAT ATC GAG CTA TGC AGT TAG CCA
arg val phe ala met asn asp gly glu arg leu arg ala ile glu val pro ala asp tyr ile glu leu cys ser  •
```

FIGURE 1: Nucleotide sequence of the 4-CBA dehalogenase coding region of the 4.5-kb DNA fragment from *Pseudomonas* sp. CBS-3. The amino acid sequences for the 4-CBA-CoA dehalogenase, 4-CBA:CoA ligase, and 4-HBA-CoA thioesterase as deduced from the sequences of ORF1, ORF2, and ORF3, respectively, are also shown. The possible ribosome binding sites are indicated with a line above the sequence.

Scheme III: A Summary of the Types of Reactions Catalyzed by the Proteins Represented in Figure 3



To gain an understanding of why this relatively diverse group of proteins might be structurally related, we first examined the similarity of the chemical reactions that they catalyze. Common to each enzyme (with the possible exception of AngR whose enzymatic activity has not yet been defined) is the catalysis of a reaction between ATP and a carboxylate substrate to form an acyl adenylate and catalysis of acyl transfer from the acyl adenylate to an acceptor. Proteins that activate their substrates by acyl phosphate formation such as glutathione synthetase (Gushima et al., 1984) or γ-glutamylcysteine synthetase (Watanabe et al., 1986) are not related (on the basis of our attempts to align them) in primary structure to the acyl adenylate forming proteins of Figure 3.

Not all proteins that catalyze the acyl adenylate formation followed by acyl transfer are related, however. In particular, the aminoacyl-tRNA synthases which catalyze adenylation of a carboxylate followed by acyl transfer to an alcohol (Burbaum et al., 1990) do not appear, on the basis of our database search, to be structurally related to the group of proteins represented in Figure 3. This observation prompted us to look for some other aspect of their chemistry which differentiated the proteins aligned in Figure 3 from these other, apparently unrelated, ligases. An obvious candidate is the chemistry of the second partial reaction that they catalyze, acyl transfer from the acyl adenylate (see Scheme III).

Common to the 4-CBA:CoA ligase, 4-coumarate:CoA ligase, fatty acid:CoA ligase, and acetate:CoA ligase is catalysis of acyl transfer to the thiol substituent of CoA (reactions I–II, Scheme III). This subgroup of proteins may be linked to a second subgroup which catalyzes acyl transfer to or from the thiol substituent of an appended phosphopantetheine arm in the thiol template-directed process that links amino acids (reactions I–III, Scheme III). For example, the EntE and EntF of the *E. coli* enterobactin synthetic apparatus catalyze the adenylation of 2,3-dihydroxybenzoate and L-serine, respectively (Rusnak et al., 1989; Reichert et al., 1992). The 2,3-dihydroxybenzoate adenylate is not released from EntE (Rusnak et al., 1989) but instead may react with the thiol moiety of the phosphopantetheine arm attached to EntF (Rusnak et al., 1991). In this manner, the 2,3-dihydroxybenzoyl group could be transferred to the amino substituent of the EntF-bound L-serine adduct. The gramicidin S synthetase and tyrocidine synthetase systems [for a recent review see Kleinkauf and von Döhren (1990)] appear to operate in a similar manner. Gramicidin S synthetase 1 and tyrocidine synthetase 1 catalyze the adenylation and racemization of L-phenylalanine, the first step common to the biosynthetic pathways leading to the two cyclic peptides gramicidin S and tyrocidine. It has been proposed that following adenylation the L-phenylalanine acyl unit is transferred to an active site cysteine and then epimerized. The multifunctional gramicidin S synthetase 2 and tyrocidine synthetase 2, which are responsible for activation and assembly of the remaining amino acids of the respective peptide, are believed to contain covalently bound phosphopantetheine. Thus, in analogy to the EntE–EntF components of the enterobactin system, one could envision acyl transfer from the cysteine linked D-phenylalanine moiety of the synthetase 1 to the phosphopantethiene thiol of synthetase 2 and hence to the amine substituent of the amino acid to which it is to couple. Yet another example of this type of chemistry is exhibited by the aminoadipyl–cysteinyl–valine synthetase (ACV synthetase) (van Liempt et al., 1989, Kleinkauf & von Döhren, 1990). This is a multienzyme complex which, in a presumed thiol template-directed process closely resembling that employed in the gramicidin S and tyrocidine synthetases, catalyzes the formation of the aminoadipate adenylate, the coupling of it to L-cysteine, and then the epimerization and coupling of L-valine.

The luciferase reaction has been described as proceeding with the formation of a luciferin–AMP adduct, a peroxyanion adduct, and finally to a light-emitting dioxetane adduct (Walsh, 1979). It has also been shown that CoA stimulates luciferase-catalyzed light emission by reacting with the enzyme-bound luciferin–AMP adduct to form the luciferin-CoA thioester (Airth et al., 1958; Rhodes & McElroy, 1958). We note that carbanion formation, which is necessary for the oxygenation step, would be more facile from the CoA thioester than from the adenylate (a point exemplified by the mechanism of epimerization of L-phenylalanine by the gramicidin S and tyrocidine synthetase 1) and speculate that CoA plays a cofactor role in the luciferase reaction.[5]

Finally, although the AngR protein is presently thought to function in the anguibactin biosynthetic pathway as a transcriptional activator (Salinas et al., 1989), its structural relatedness to the proteins shown in Figure 3 might suggest otherwise. The chemical steps leading from 2,3-dihydroxybenzoate and cysteine to the anguibactin structure (Jalal et al., 1989) have not yet been determined. However, one can speculate that the thiozole ring of the anguibactin is constructed by adenylation of the carboxyl substituent of 2,3-dihydroxybenzoate followed by acyl transfer to the cysteine thiol group and then ring closure to the imine. It is possible that the AngR was conscripted for its regulatory role from a catalytically active ancestor.[6]

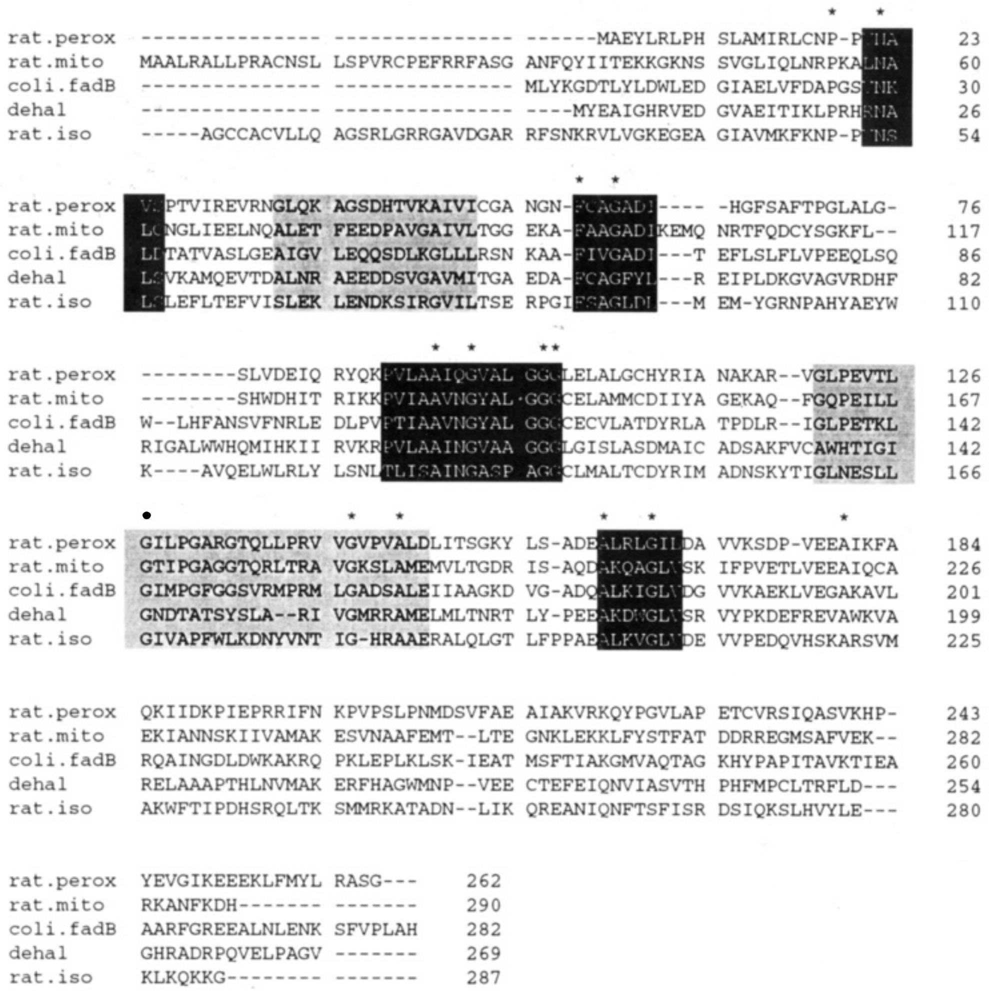
The data currently available on the chemical pathways catalyzed by the proteins whose amino acid sequences we have aligned in Figure 3 are, in many cases, severely limited. Nevertheless, these data do raise the possibility that this group of proteins are functionally related not only by catalysis of carboxyl group activation by adenyl transfer from ATP but [as first noted by Toh (1990) for a subgroup of the proteins shown in Figure 3] also by subsequent acyl transfer to a thiol as an intermediate or final step in the chemical pathway. In the case of the multienzyme synthetases (enterobactin, gramicidin S, tyrocidine, and ACV synthetase) the thioester is utilized as part of the thiol template-coupling strategy. The conversion of the carboxylate substrate to the thioester occurs "in house" so to speak. The carboxylate:CoA ligases (4-CBA:CoA, 4-coumarate:CoA, fatty acid:CoA, and acetate: CoA ligases), on the other hand, could be viewed as detached protein units functioning in concert with the enzymes associated with their respective biodegradative or biosynthetic pathways.

Having explored the similarities among the proteins listed in Table I in relation to the chemistry that they catalyze, we now turn to a more detailed examination of the similarities among their primary structures.[7] The sequences most like each other have been grouped at the top of the alignment with the less related sequences grouped at the bottom of the alignment. The overall structural relatedness of these proteins is most apparent from the presence of the particularly highly conserved sequence motif (T,S)(S,G)G(T,S)(T,E)G(L,X)-PK(G–) (residues 161–170 in the 4-CBA:CoA ligase sequence) in each of the protein sequences (shaded black in Figure 3). From the other black-shaded regions in Figure 3, it can be seen

[5] Rhodes and McElroy (1958) interpreted the CoA enhancement effect as resulting from the liberation of enzyme (from the luciferin–AMP adduct) to recycle for further light production. The finding that cysteine would not substitute for CoA (Rhodes & McElroy, 1958) suggests to us a specific role for CoA in luciferase catalysis.

[6] Concription of a protein for a new and unrelated function has precedent in the evolutionary mechanism that has been proposed for the pair of homologous proteins argininosuccinate lyase and δ-crystallin (Yeh et al., 1988; Piatigorsky et al., 1988).

[7] One result of generating an alignment of this size is that new information about the conservation of structural features in all of the proteins becomes more apparent. For example, there has been considerable speculation about the possible "essential" role of a cysteine (Hori et al., 1989) which appears at residue 376 in the gramicidin S synthetase 1 sequence. This cysteine has been identified as conserved in the gramicidin S and tyrocidine synthetase 1 sequences because it may be essential to the thiol template mechanism of peptide synthesis. It is also conserved in the coumerate:CoA ligase and luciferase sequences (Becker-Andre et al., 1991). Others note, however, that this Cys is not conserved in the ACV synthetase sequences, even though these proteins are responsible for synthesizing a related peptide antibiotic (MacCabe et al., 1991), or in the enterobactin F sequence (Rusnak et al., 1991). Our alignment shows that this Cys is not conserved in several other sequences in the alignment as well and is, in fact, less conserved than a number of residues in the nearby vicinity. Another point of interest concerns the observation that the AngR protein exhibits a helix–turn–helix motif showing striking homology with prokaryotic DNA-binding proteins such as λ and P22 Cro (Farrell et al., 1990). The residues in question, 873–894 of the AngR sequence, appear to be in a region of generally low similarity between the AngR and the rest of the sequences. This result is consistent with the fact that none of the other proteins are as likely to bind DNA.



```
                                                                    *  *
rat.perox  --------------- --------------- ------MAEYLRLPH SLAMIRLCNP-PVNA   23
rat.mito   MAALRALLPRACNSL LSPVRCPEFRRFASG ANFQYIITEKKGKNS SVGLIQLNRPKALNA   60
coli.fadB  --------------- --------------- MLYKGDTLYLDWLED GIAELVFDAPGSVNK   30
dehal      --------------- --------------- ----MYEAIGHRVED GVAEITIKLPRHRNA   26
rat.iso    -----AGCCACVLLQ AGSRLGRRGAVDGAR RFSNKRVLVGKEGEA GIAVMKFKNP-PVNS   54

                                               *  *
rat.perox  VSPTVIREVRNGLQK AGSDHTVKAIVICGA NGN-FCAGADI---- -HGFSAFTPGLALG-   76
rat.mito   LCNGLIEELNQALET FEEDPAVGAIVLTGG EKA-FAAGADIKEMQ NRTFQDCYSGKFL--  117
coli.fadB  LDTATVASLGEAIGV LEQQSDLKGLLLRSN KAA-FIVGADI---T EFLSLFLVPEEQLSQ   86
dehal      LSVKAMQEVTDALNR AEEDDSVGAVMITGA EDA-FCAGFYL---R EIPLDKGVAGVRDHF   82
rat.iso    LSLEFLTEFVISLEK LENDKSIRGVILTSE RPGIFSAGLDL---M EM-YGRNPAHYAEYW  110

                                   *  *     **
rat.perox  --------SLVDEIQ RYQKPVLAAIQGVAL GGGLELALGCHYRIA NAKAR--VGLPEVTL  126
rat.mito   --------SHWDHIT RIKKPVIAAVNGYAL GGGCELAMMCDIIYA GEKAQ--FGQPEILL  167
coli.fadB  W--LHFANSVFNRLE DLPVPTIAAVNGYAL GGGCECVLATDYRLA TPDLR--IGLPETKL  142
dehal      RIGALWWHQMIHKII RVKRPVLAAINGVAA GGGLGISLASDMAIC ADSAKFVCAWHTIGI  142
rat.iso    K----AVQELWLRLY LSNLTLISAINGASP AGGCLMALTCDYRIM ADNSKYTIGLNESLL  166

           ●                *   *                *   *               *
rat.perox  GILPGARGTQLLPRV VGVPVALDLITSGKY LS-ADEALRLGILDA VVKSDP-VEEAIKFA  184
rat.mito   GTIPGAGGTQRLTRA VGKSLAMEMVLTGDR IS-AQDAKQAGLVSK IFPVETLVEEAIQCA  226
coli.fadB  GIMPGFGGSVRMPRM LGADSALEIIAAGKD VG-ADQALKIGLVDG VVKAEKLVEGAKAVL  201
dehal      GNDTATSYSLA--RI VGMRRAMELMLTNRT LY-PEEAKDMGLVSR VYPKDEFREVAWKVA  199
rat.iso    GIVAPFWLKDNYVNT IG-HRAAERALQLGT LFPPAEALKVGLVDE VVPEDQVHSKARSVM  225

rat.perox  QKIIDKPIEPRRIFN KPVPSLPNMDSVFAE AIAKVRKQYPGVLAP ETCVRSIQASVKHP-  243
rat.mito   EKIANNSKIIVAMAK ESVNAAFEMT--LTE GNKLEKKLFYSTFAT DDRREGMSAFVEK--  282
coli.fadB  RQAINGDLDWKAKRQ PKLEPLKLSK-IEAT MSFTIAKGMVAQTAG KHYPAPITAVKTIEA  260
dehal      RELAAAPTHLNVMAK ERFHAGWMNP--VEE CTEFEIQNVIASVTH PHFMPCLTRFLD---  254
rat.iso    AKWFTIPDHSRQLTK SMMRKATADN--LIK QREANIQNFTSFISR DSIQKSLHVYLE---  280

rat.perox  YEVGIKEEEKLFMYL RASG---   262
rat.mito   RKANFKDH------- -------   290
coli.fadB  AARFGREEALNLENK SFVPLAH   282
dehal      GHRADRPQVELPAGV -------   269
rat.iso    KLKQKKG-------- -------   287
```

FIGURE 2: Linear alignment of the amino acid sequences of the 2-enoyl-CoA hydratase domain of the trifunctional rat liver peroxisomal enzyme (rat.perox) (Osumi et al., 1985), the monofunctional rat liver mitochondrial 2-enoyl-CoA hydratase (rat.mito) (Minami-Ishii et al., 1989), the 2-enoyl-CoA hydratase domain of the multifunctional *E. coli* enzyme (coli.fad B) (Dirusso et al., 1990; Nakahigashi & Inokuchi, 1990; Yang et al., 1991), and the 4-CBA-CoA dehalogenase (dehal) and the monofunctional rat liver mitochondrial Δ^3-*cis*-Δ^2-*trans*-enoyl-CoA isomerase (rat.iso) (Palosaari et al., 1991) with amino acid numbering shown in the righthand margin. Residues conserved throughout all five sequences are indicated by an asterisk while regions of highest sequence identity are shaded black. Other regions of high sequence similarity are shaded grey.

that additional sequence relationship extends beyond the highly conserved motif mentioned above. Statistical analysis of these sequences suggests "probable relationship" between each protein and at least one other in the set. Both the statistical analysis and examination of the alignment itself suggest that some of these sequences fall into subgroups which exhibit similarities which distinguish the members of that group. These regions are shaded dark grey in Figure 3 and are associated with the proteins involved in antibiotic synthesis, the gramicidin S and tyrocidine synthetases and ACV synthetase, the siderophore synthetic proteins EntF and AngR, and finally, α-aminoadipate reductase. Thus, these proteins appear to be more related to each other than to the other sequences. As might be expected, the two luciferase sequences are more related to each other than to the other proteins. These luciferase sequences, along with the coumarate:CoA ligase and the fatty acid:CoA ligase sequence, form another, less well-defined, subset (in which the regions of interest are shaded light grey). The protein whose sequence is reported in this work, the 4-CBA:CoA ligase of the dehalogenase system, does not appear to belong primarily to either of these subgroups, but it appears to be most related, overall, to the EntE sequence. Taken together, these results can be interpreted to suggest that all of these proteins evolved by divergence from a primitive ancestor rather than being related by convergence to a single important primary motif. The low level of overall similarity among these protein families (below 30% for most protein pairs) suggests that they represent distantly diverged gene duplications.

The extent to which the (T,S)(S,G)G(T,S)(T,E)G(L,X)-PK(G–) sequence motif is conserved among all of the proteins listed in Table I is so striking that it must surely be a functional motif. Neither this motif nor the sequences associated with the other conserved regions identified in Figure 3 correspond to known functional motifs, however. In fact, we nor others (Masuda et al., 1989; Staab et al., 1989; Rusnak et al., 1991; Hori et al., 1991) could find evidence of a phosphate-binding loop motif (Saraste et al., 1990) in any of these proteins. Thus, the sequence regions conserved among this large family of proteins, and in particular, the (T,S)(S,G)G(T,S)(T,E)G-(L,X)PK(G–) sequence motif, represent a newly discovered functional motif which we suggest is likely to be related to acyl adenylate formation and possibly, thioester formation. Site-directed mutagenesis experiments with the 4-CBA:CoA ligase are underway to examine the functional role of this new motif.

Search for Sequence Identity with the 16-kDa Polypeptide, 4-HBA-CoA Thioesterase. A comparison of the amino acid sequence of the 16-kDa polypeptide encoded by ORF3 of Figure 1 with the protein sequences contained in the PIR and EMBL databanks failed to identify any proteins which share

```
                                                    *    *  **
gram.syn.1  LVHLIHNIQFNGQVE IFEEDTIKIREGTNL HVPSKSTDLAYVIYT SGTTGKPKGTMLEHK GISNLKVFFENSLNV TEKDRIGQ--FASIS   233
tyr.syn.1   VSQLVHDVGYSGEVV VLDEEQLDARETANL HQPSKPTDLAYVIYT SGTTGKPKGTMLEHK GIAICNPFSKIRLAS PSKTGSG---FLPAC   220
gram.syn.2  --HLKDKFAFTKETI VIEDPSISHELTEEI DYINESEDLFYIIYT SGTTGKPKGVMLEHK NIVNLLHFTFEKTNI NFSDKVLQ--YTTCS   657
acv.syn.B   RIKGMAASGTLLYPS VLPANPDSKWSVSNP SPLSRSTDLAYIIYT SGTTGRPKGVTVEHH GVVNLQVSLSKVFGL RDTDDEVILSFSNYV  1612
ent.syn.F   --DQLPRFSDVPNLT SLCYNAPLTPQGSAP LQLSQPHHTAYIIFT SGSTGRPKGVMVGQT AIVNRLLWMQNHYPL TGEDVVA--QKTPCS   646
ang.R       --SDSKNSPSNDLFF FLDWQTAIKSEPMRS PQDVAPSQPAYIIYT SGSTGTPKGVVISHQ GALNTCIAINRRYQI GKNDRVL--ALSALH   644
adip.red    IQENGTIEGGKLDNG EDVLAPYDHYKDTRT GVVVGPDSNPTLSFT SGSEGIPKGVLGRHF SLAYYFNWMSKRFNL TENDKFT--MLSGIA   464
beet.luci   NIHGCESLPNFISR- --Y--SKGNIANFKP LHYDPVEQVAAILCS SGTTGLPKGVMQTHR NVCVRLIHALDPRVG TGLI-PGVTVLVYLP   239
fire.luci   DYQGFQSMYTFVTS- --HLPPGFNEYDFVP ESFDRDKTIALIMNS SGSTGLPKGVALPHR TACVRFSHARDPIFG NQII-PDTAILSVVP   242
coum.lig    DCLHFSKLMEADE-- -----------SEMP EVVINSDDVVALPYS SGTTGLPKGVMLTHK GLVTSVAQQVDGDNP NLYMHSEDVMICILP   235
f.a.lig     NDLVERGQKCGVEII GLKALEDLGRVNRTK PKPPEPEDLAIICFT SGTTGNPKGAMVTHQ NIMNDCSGFIKATES AFIASPEDVLISFLP   322
4CBA.lig    GSRVPE--------- --------VAGKSLP PPPCEPEQASFVFYT SGTTGLPKGAVIPQR AAESRVLFMATQAGL RHGSHNVVLGLMPLY   206
ent.syn.E   EHNLQDA-------- ------INHPAEDFT ATPSPADEVAYFQLS GGTTGTPK--LIPRT HNDYYYSVRRSVEIC QFTQQTRYLCAIPAA   233
acet.lig    PIISMTPGRDLWW-- ---HEEVEKYPAYYT PVAMASEDPLFLLYT SGSTGKPK--GVAHS TGGYLLRAMTGKYVF DIHDGDRYFCGGDVG   286

gram.syn.1  FDASVWEMFMALLTG ASLYIILKDTINDFV KFEQYINQKEITVIT LPPTYVVHLDPER-- --ILSIQTLITAGSA TSPSLV--NKWK---   314
tyr.syn.1   RSTHPFGKCSWLCCL APRVHPSKQTIHDFA AFEHYLSENELTIIT LPPTYLTHLTPER-- --ITSLRIMITAGSA SSAPLV--NKWK---   301
gram.syn.2  FDVCYQEIFSTLLSG GQLYLIRKETQRDVE QLFDLVKRENIEVLS FPVAFLKFIFNERE- -FINRFPTCVKHIIT AGEQLVVNNEFKR-Y   744
acv.syn.B   FDHFVEQMTDAILNG QTLLVLNDGMRGDKE RLYRYIEKNRVTYLS GTPSVVSMYEFSR-- --FKDHLRRVDCVGE AFSEPV-FDKIR---  1694
ent.syn.F   FDVSVWEFFWPFIAG AKLVMAEPEAHRDPL AMQQFFAEYGVTTTH FVPSMLAAFVASLTP QTARQSCATLKQVFC SGEALP-ADLCRE-W   734
ang.R       FDLSVYDIFGLLSAG GTIVLVSELERRDPI AWCQAIEEHNVTMWN SVPALFDMLLTYA-T CFNSIAPSKLRLTML SGDWIG-LDLPQRYR   732
adip.red    HDPIQRDMFTPLFLG AQLYVPTQDDIGTPG RLAEWMSKYGCTVTH LTPAMGQLLTAQA-- --TTPFPKLHHAFFV GDILTKR-DCLR-LQ   548
beet.luci   FFHAFGFSINLGYFM -VGLRVIMLRRFDQE AFLKAIQDYE--VRS VINVPAII----LFL SKSPL---------- ---------------   297
fire.luci   FHHGFGMFTTLGYLI -CGFRVVLMYRFEEE LFLRSLQDYK--IQS ALLVPTLF----SFF AKSTL---------- ---------------   300
coum.lig    LFHIYSLNAVLCCGL RAGVTILIMQKFDIV PFLELIQKYK--VTI GPFVPPIV----LAI AKSPV---------- ---------------   294
f.a.lig     LAHMFETVVECVMLC -HGAKIGFF-QGDIR LLMDDLKVLQ--PTI FPVVPRLL----NRM FDRIFGQANTSVKRW LLDFASKRKEAELRS   404
4CBA.lig    HTIGFFAVLVAAMAF DGTYVVV--EEFDAG NVLKLIERER--VTA MFATPTHL----DAL TTAVE---------- ---------------   263
ent.syn.E   HNYAMSSPGSLGVFL AGGTVVLA-ADPSAT LCFPLIEKHQVNVTA LVPPAVSL----WLQ ALIEG---------- ---------------   293
acet.lig    WITGPHYVLSAPLLL GVSTVVFEGTPPTNS PYWDIIEEHK--VTQ FSVAPTALRLLKRAG DHHVR---------- ---------------   349

                                                                  *
gram.syn.1  ---EKVTYI------ --------------- --------------- ---------MAYGPT ETTICATTWVA---- ----TKETTGHSVPI   348
tyr.syn.1   ---DKLRYI------ --------------- --------------- ---------MAYGPT ETSICATIWEAP--- ----SNQLSVQSVPI   336
gram.syn.2  LHEHNVHLH------ --------------- --------------- ---------MHYGPS ETHVVTTYTIN---- ----PEAEIPELPPI   781
acv.syn.B   -ETFHGLVI------ --------------- --------------- ---------MGYGPT EVSITTHKRLY---- ----PFPERRMDKSI  1730
ent.syn.F   QQLTGAPLH------ --------------- --------------- ---------ILYGPT EAAVDVSWYPAF--- -GEELAQVRGSSVPI   775
ang.R       NYRVDGQFI------ --------------- --------------- ---------AMGGAT EASIWSNVFDV---- ---EKVPMEWRSIPY   770
adip.red    TLAENCRIV------ --------------- --------------- ---------MMYGTT ETQRAVSYFEVKSKN DDPNFLKKLKDVMPA   593
beet.luci   --------------- -VDKYDLSSLRELCC GAAPLAKEVAEIAVK RLNPLPGIRCGFGLT EST--S-----ANIH SLRDEFKSGSLGRVT   364
fire.luci   --------------- -IDKYDLSNLHEIAS GGAPLSKEVGE-AVA KRFHLPGIRQGYGLT ETT--S-----AILI TPEGDDKPGAVGKVV   366
coum.lig    --------------- -VDKYDLSSVRTVMS GAAPLGKELED-AVR AKFPNAKLGQGYGMT EAG--PVLAMCLAFA KEPYEIKSGACGTVV   365
f.a.lig     GIVRNNSLWDKLIFH KIQSSLGGKVRLMIT GAAPVSATVL--TFL RAALGCQFYEGYGQT ECT--A-----GCCL SLPGDWTAGHVGAPM   485
4CBA.lig    --------------- -QAGARLESLEHVTF AGATMPDTVL--ERV NRFIPGEKVIIYGTT EAM----NSLYMRAV RIAGTVMRPGFFSEV   331
ent.syn.E   --------------- -ESRAQLASLKLLQV GGARLSATLA--ARI PAEIGCQLQQVFGMA EGLV-NYTRLDDSAE KIIHTQGYPMCPDDE   364
acet.lig    --------------- -NEMKHLRVLGSVGE PSAAEVWKWY--YDV VGKAAAQICDTYWQT ETGSNVITPLAGVTP TKPGSASFPFFGIEP   421

gram.syn.1  GAPIQNTQIYIV--D ENLQLKSVGEAGELC IGGEGLARGYWKRPE LTSQKFVDNPFVPG- -------------E- -------KLYKTGDQ   414
tyr.syn.1   GKPIQNTHIYIV--N EDLQLLPTADEGELC IGGVGLARGYWNRPD LTAEKFVDNPFVPG- -------------E- -------KMYRTGDL   402
gram.syn.2  GKPISNTWIYIL--D QEQQLPQGIVGELY ISGANVGRGYLNNNE LTAEKFFADPFRPN- -------------E- -------RMYRTGDL   847
acv.syn.B   GQQVHNSTSYVL--N EDMKRTPIGAVGELY LGGEGVVRGYHNRAD VTAERFIPNPFQSE- -------------ED KREGRNSRLYKTGDL  1804
ent.syn.F   GYPVWNTGLRIL--D AMMHPVPPGVAGDLY LTGIQLAQGYLGRPD LTASRFIADPFAPG- -------------E- -------RMYRTGDV   841
ang.R       GYPLPRQQYRVV--D DLGRDCPDWVAGELN IGGDGIALGYFDDEL KTQAQFLHIDGHA-- --------------- --------WYRTGDM   833
adip.red    GKGMLNVQLLVVNRN DRTQICGIGEIGEIY VRAGGLAEGYRGLPE LNKEKFVNNWFVEKD HWNYLDKDNGEPWRQ FWLGPRDRLYRTGDL   683
beet.luci   PLMAAKIADRETGKA LGPNQV-----GELC IKGPMVSKGYVNNVE ATKEAIDDDGWLHSG DFGYYDEDEHFYVVD R-YKELIKY-KGSQV   447
fire.luci   PFFEAKVVDLDTGKT LGVNQR-----GELC VRGPMIMSGYVNNPE ATNALIDKDGWLHSG DIAYWDEDEHFFIVD R-LKSLIKY-KGYQV   449
coum.lig    RNAEMKIVDPETNAS LPRNQR-----GEIC IRGDQIMKGYLNDPE STRTTIDEEGWLHTG DIGFIDDDDELFIVD R-LKEIIKY-KGFQV   448
f.a.lig     PCNYIKLVDVEDMNY QAAKGE-----GEVC VKGANVFKGYLKDPA RTAEALDKDGWLHTG DIGKWLPNGTLKIID R-KKHIFKLAQGEYI   569
4CBA.lig    RIVRVGGDVDDGCPT V----------KRAS WRWRRRMRPFQATLT NLRLLQKSFRKAGTG RAICVR----DGSGN IVVLGRVDDMIIS--   405
ent.syn.E   VWVAECRRKSTAARE V----------GELH TRGPYTFRGYYKSPQ HNASAFDANGFYCSG DLISI-----DPEGY ITVQGREKDQINR--   437
acet.lig    ALVDPVTGEEIRGND V----------EGVL AFKQPWPSMARTVWG AHKRYMETYLHVYKG YYFTGDGAARDHEGF YWIRGR-VDRVNV--   498

gram.syn.1  ARWL--PDGNIEYLG RIDNQVKIRGHRVEL EEVESILLKHMYISE TAV------------ --------------- SVHKDHQEQPYLCAI   475
tyr.syn.1   AKWL--TDGTIEYLG RIDHQVKIRGHRIEL GEIESVLLAHEHITE AVV------------ --------------- IAREDQHAGQYLCAY   463
gram.syn.2  ARW[illegible] [illegible]ADHQV[illegible]GHRIEL GEIEAQLLNCKGVKE AVV------------ --------------- IDKADDKGGKYLCAY   908
acv.syn.B   VRW[illegible] [illegible]GLRIEL GEIEAILSSYHGIKQ SVVI----------- --------------- AKDCREGAQKFLVGY  1868
ent.syn.F   ARW[illegible] [illegible]GQRIEL GEIDRVMQALPDVEQ AVTH----------- ----------ACVIN QAAATGGDARQLVGY   908
ang.R       GGK[illegible] [illegible]GYRIEL GEIEVALNNIPGVQR AVA------------ -----------IAVG NKDKTLAAFIVMDSE   898
adip.red    [illegible] [illegible]GFRIEL GEIDTHISQHPLVRE NITLVRKNADNEPTL ITFMVPRFDKPDDLS KFQSDVPKEVETDPI   771
beet.luci   APAELEEILLKNPCI RDVAVVG---IPDLE AGELPSAFVVI-QPG KEITAKEV-YDYLAE RVSHTKYLRGGVRFV DSIPRNVTGKITRKE   532
fire.luci   APAELESILLQHPNI FDAGVAG---LPDDD AGELPAAVVVL-EHG KTMTEKEI-VDYVAS QVTTAKKLRGGVVFV DEVPKGLTGKLDARK   534
coum.lig    APAELEALLLTHPTI SDAAVVP---MIDEK AGEVPVAFVVR-TNG FTTTEEEI-KQFVSK QVVFYKRI-FRVFFV DAIPKSPSGKILRKD   532
f.a.lig     APEKIENIYLRSEAV AQVFVHGESLQAFLI AIVVPDVEILP-SWA QKRGFQGS-FEELCR NKDINKAI---LEDM VKLGKNAGLKPFEQ-   653
4CBA.lig    GGENIHPSEVERILA AAPGVAEVVVIG--V KDERWGQSVVACVVL QP-GASASAERLD-- AFCRASALADFKRPR RYVFLDELPKSAVNK   490
ent.syn.E   GGEKIAAEEIENLLL RHPAVIYAALVS--M EDELMGEKSCAYLVV K---EPLRAVQVR-- RFLREQGIAEFKLPD RVECVDSLPLTAVGK   520
acet.lig    SGHRLSTAEIEAALI EHHSIAEAAVVG--V ADELTGQAVNAFVAV KEGTQINDALRKESP SLQVRRSIGPFAAPK AIYIVPDLPKTLSGK   586
```

FIGURE 3: Linear alignment of the amino acid sequences of *Bacillus brevis* Nagano gramicidin S synthetase 1 (of ram.syn. 1) (Hori et al., 1989), *B. brevis* tyrocidine synthetase (tyrsyn. 1) (Weckermann et al., 1988), *B. brevis* Nagano gramicidin S synthetase 2 (gram.syn 2) (Hori et al., 1991), *Penicillium chryosogenum* α-aminoadipyl–cysteinyl–valine (ACV) synthetase domain B (acv.syn. B) (Diez et al., 1990), *E. coli* enterobactin synthetase component F (ent.syn. F) (Rusnak et al., 1991), *Vibrio anguillarum* anguibactin synthetase component R (ang.R) (Farrell et al., 1990), *Saccharomyces cerevisiae* α-aminoadipate reductase (adip.red) (Morris et al., 1991), *Pyrophorus plagiophthalamus* (green light emitting) photinus–luciferin 4-monooxygenase (beet.luci) (Wood et al., 1989), *Photinus pyralis* photinus–luciferin 4-monooxygenase (fire.luci) (de Wet et al., 1987), *Petroselinum crispum* 4-coumarate:CoA ligase (coum.lig) (Lozoya et al., 1988), *Rattus norvegicus* long-chain fatty acid:CoA ligase (f.a.lig) (Suzuki et al., 1990), *E. coli* enterobactin synthetase component E (ent.syn. E) (Staab et al., 1989), and *Neurospora crassa* acetate:CoA ligase (acet.lig) (Connerton et al., 1990) with amino acid numbering shown in the righthand margin. Residues conserved throughout all 14 sequences are indicated by an asterisk while regions of high sequence identity among all or most of the sequences are shaded in black. Regions of high sequence similarity particular to the subgroup gram.syn. 1, tyr.syn.1, gram.syn. 2, acv.syn.B., ent.syn.F., ang.R., and adip.red are shaded in dark grey while regions of high sequence similarity particular to the subgroup beet.luci, fire.luci, coum.lig, and f.a.lig are shaded in light grey.

with it a significant level of identity. Particular scrutiny was applied to the sequences representing the thioesterases and acyl carrier proteins of the fatty acid synthetic apparatus and to the thioesterase domains of the gramicidin S, ACV, and anguibactin synthetic complexes. An alignment was generated of 12 known thioesterase sequences (data not shown) which revealed structural relatedness between the thioesterases of the fatty acid biosynthetic pathway and between the thioesterases of the antibiotic (gramacidin S, ACV) and siderophore (anguibactin) synthetic pathways. While these two families of thioesterases appear to be remotely related by sequence, neither group is related to the 4-HBA-CoA thioesterase of the 4-CBA dehalogenation pathway. Furthermore, the 4-HBA-CoA thioesterase sequence does not contain the functional motif GXSXG, found in lipases, thioesterases, and the acetyltransferase subunit of fatty acid synthetase (Mikkelsen et al., 1985; Kräetzschmar et al., 1989; Brady et al., 1990; Winkler et al., 1990). The lack of a relationship between the 4-HBA-CoA thioesterase and the other thioesterases we examined suggests that the dehalogenase pathway enzyme may have been recruited from a different pathway, one which may become evident as new thioesterase amino acid sequences are discovered.

Conclusion. The chemical strategy that we have seen unfold for the biodegradation of 4-CBA in *Pseudomonas* sp. CBS-3 involves the coupling of three catabolic pathways. The first pathway converts the 4-CBA to 4-HBA in a three-step process. The 4-HBA thus formed is then oxidized to protochuate and then to carboxymuconate via the ortho-cleavage pathway (Ornston, 1990). The β-ketoadipate which is ultimately derived is converted via β-oxidation in the β-ketoadipate pathway to succinyl-CoA and acetyl-CoA. Thus, the dehalogenase pathway described herein allows this particular pseudomonad to utilize an unusual supplementary carbon source, 4-CBA, metabolizing it to a form that can enter a conventional aromatic metabolizing pathway.

From the present study, we have determined some of the relatives of the 4-CBA dehalogenation pathway enzymes. Our findings have been limited by the low degree of overall sequence identity existing between the related proteins and by the small proportion of protein sequences which have been deduced. While the 4-CBA:CoA ligase was linked with several distinctly different protein families, the 4-CBA-CoA dehalogenase was linked with only a single family of proteins (note that sequences of enoyl-hydratases from β-oxidation pathways other than the fatty acid pathway are not yet available) and the 4-HBA-CoA thioesterase was linked with none.

The reaction steps constituting the 4-CBA dehalogenation pathway (viz., CoA thioesterification of a carboxylate, hydration of the carbon–carbon double bond of a conjugated enoyl-CoA thioester, and hydrolysis of a CoA thioester) (see Scheme I) parallel those of the fatty acid, the amino acid, and most particularly, the 4-coumarate β-oxidation pathways. From this perspective, the observation that the 4-CBA-CoA dehalogenase is related to a family of 2-enoyl-CoA hydratases of the fatty acid β-oxidation pathway and that the 4-CBA:CoA ligase is related to a large group of ligases which includes 4-coumarate:CoA ligase and fatty acid:CoA ligase is most reasonable. The 4-coumarate:CoA ligase catalyzes the first step of the β-oxidation pathway leading from 4-coumarate to 4-HBA in plants (Goodwin & Mercer, 1983) while the fatty acid:CoA ligase initiates fatty acid β-oxidation in all organisms. While it is possible that the three genes encoding the 4-CBA → 4-HBA-converting enzymes were recruited from different gene clusters encoding enzymes of separate metabolic pathways, it seems more probable that the three genes were derived from a single gene locus encoding the enzymes of a β-oxidation pathway. As new sequences are determined, it should be possible to trace the connections to such an ancestral pathway in more detail.

Acknowledgments

We thank Drs. Reichert and Walsh for providing us with their unpublished manuscript. We also thank Ms. Leslie Taylor for assistance with the sequence analysis software.

Supplementary Material Available

Figure 1A showing the DNA sequencing strategy used for the 3.0-kb *Sal*I–*Sal*I fragment in pUC19 and Figure 1B showing the strategy used for the sequencing of the 1.6-kb *Sal*I–*Sma*I DNA fragment (2 pages). Ordering information is given on any current masthead page.

References

Abramowicz, D. A. (1990) *Crit. Rev. Biotechnol. 10*, 241.

Airth, R. L., Rhodes, W. C., & McElroy, W. D. (1985) *Biochim. Biophys. Acta 27*, 519.

Altschul, S. F., & Erickson, B. W. (1986) *Bull. Math. Biol. 48*, 603.

Bahnson, B. J., & Anderson, U. E. (1991) *Biochemistry 30*, 5894.

Becker-André, M., Schulze-Lefert, P., & Hahlbrock, K. (1991) *J. Biol. Chem. 266*, 8551.

Brady, L., Brzozowski, A. M., Derewenda, Z. S., Dodson, E., Dodson, G., Tolley, S., Turkenburg, J. P., Christiansen, C., Huge-Jensen, B., Norskov, L., Thim, L., & Menger, U. (1990) *Nature 343*, 767.

Burbaum, J. J., Starzyk, R. M., & Schimmel, P. (1990) *Proteins: Struct., Funct., Genet. 7*, 99.

Chang, K.-H., Liang, P.-H., Beck, W., Scholten, J. D., & Dunaway-Mariano, D. (1992) *Biochemistry* (following paper in this issue).

Commandeur, L. C. M., & Parsons, J. R. (1990) *Biodegradation 1*, 207.

Connerton, I. F., Fincham, J. R. S., Sandeman, R. A., & Hynes, M. J. (1990) *Mol. Microbiol. 4*, 451.

deWet, J. R., Wood, K. V., DeLuca, M., Helinski, D. R., & Subramani, S. (1987) *Mol. Cell. Biol. 7*, 725.

Diéz, B., Gutiérrez, S., Barredo, J. L., van Solingen, P., van der Voort, L. H. N., & Martín, J. F. (1990) *J. Biol. Chem. 265*, 16358.

Dirusso, C. C. (1990) *J. Bacteriol. 172*, 6459.

Eurler-Bertam, S., & Stoffel, W. (1990) *Biol. Chem. Hoppe-Seyler 371*, 603.

Farrell, D. H., Mikesell, P., Actis, L. A., & Crosa, J. H. (1990) *Gene 86*, 45.

George, D. G., Barker, W. C., & Hunt, L. T. (1986) *Nucleic Acids Res. 14*, 11.

Goodwin, T. W., & Mercer, E. I. (1983) *Introduction to Plant Biochemistry*, 2nd ed., Pergamon Press LTD, U.K.

Gushima, H., Yasuda, S., Soeda, E., Yokota, M., Kondo, M., & Kimura, A. (1984) *Nucleic Acids Res. 12*, 9299.

Gutiérrez, S., Diéz, B., Montenegro, E., & Martín, J. F. (1991) *J. Bacteriol. 173*, 2354.

Hamm, G. H., & Cameron, G. N. (1986) *Nucleic Acids Res. 14*, 5.

Hawkinson, D. C., Eames, T. C., & Pollack, R. M. (1991) *Biochemistry 30*, 10849.

Henikoff, S. (1984) *Gene 28*, 351.

Hori, K., Yamamoto, Y., Minetoki, T. Kurotsu, T., Kanda, M., Miura, S., Okamura, K., Furuyama, J., & Saito, Y. (1989) *J. Biochem. 106*, 639.

Hori, K., Yamamoto, Y., Tokita, K., Saito, F., Kurotsu, T., Kanda, M., Okamura, K., Furuyama, J., & Saito, Y. (1991) *J. Biochem. 110*, 111.

Ishii, N., Hijkata, M., Osumi, T., & Hashimoto, T. (1987) *J. Biol. Chem. 262*, 8144.

Jalal, M. A., Hossain, M. B., van der Helm, D., Sanders-Loehr, J., Actis, L. A., & Crosa, J. H. (1989) *J. Am. Chem. Soc. 111*, 292.

Keil, H., Klages, U., & Lingens, F. (1981) *FEMS Microbiol. Lett. 10*, 213.

Klages, U., Markus, A., & Lingens, F. (1981) *J. Bacteriol. 146*, 164.

Klages, U., Krauss, S., & Lingens, F. (1983) *Hoppe-Seyler's Z. Physiol. Chem. 364*, 529.

Kleinkauf, H., & von Döhren, H. (1990) *Eur. J. Biochem. 192*, 3680.

Kräetzschmar, J., Krause, M., & Marahiel, M. A. (1989) *J. Bacteriol. 171*, 5422.

Kuliopulos, A., Mullen, G. P., Xue, L., & Mildvan, A. S. (1991) *Biochemistry 30*, 3169.

Lawrence, C. B., & Goldman, D. A. (1988) *Comput. Appl. Biosci. 4*, 25.

Lozoya, E., Hoffmann, H., Douglas, C., Schulz, W., Scheel, D., & Hahlbrock, K. (1988) *Eur. J. Biochem. 176*, 661.

MacCabe, A. P., van Liempt, H., Palissa, H., Unkles, S. E., Riach, M. B. R., Pfeifer, E., von Döhren, H., & Kinghorn, J. R. (1991) *J. Biol. Chem. 266*, 12646.

Markus, A., Kloges, U., Krauss, S., & Lingens, F. (1984) *J. Bacteriol. 160*, 618.

Masuda, T., Tatsumi, H., & Nakano, E. (1989) *Gene 77*, 265.

Mikkelson, J., Hojrup, P., Rasmussen, M. M., Roepstorff, P., & Knudsen, J. (1985) *Biochem. J. 227*, 21.

Minami-Ishii, N., Taketani, S., Osumi, T., & Hashimoto, T., (1989) *Eur. J. Biochem. 185*, 73.

Mirande, M., & Waller, J. P. (1988) *J. Biol. Chem. 263*, 18443.

Morris, M. E., & Jinks-Robertson, S. (1991) *Gene 98*, 141.

Müller, R., Thiele, J., Klages, U., & Lingens, F. (1984) *Biochem. Biophys. Res. Commun. 124*, 178.

Nakahigashi, K., & Inokuchi, H. (1990) *Nucleic Acids Res. 18*, 4937.

Nuttley, W. M., Aitchison, J. D., & Rachubinski, R. A. (1988) *Gene 69*, 171.

Ornston, L. N. (1990) in *Proceedings of the Genetics of Industrial Microorganisms* (Heslot, H., Davies, J., Florent, J., Bobichon, L., Duran, G., & Penasse, L., Eds.) Vol. 2, pp 1061, Microbiologie, Sociĕte Francaise de Microbiologie, Strasbourg.

Osumi, T., Ishii, N., Hijakata, M., Kamijo, K., Ozasa, H., Furuta, S., Miyazawa, S., Kondo, K., Inoue, K., Kagamiyama, H., & Hashimoto, T. (1985) *J. Biol. Chem. 260*, 8905.

Palosaari, P. M., & Hiltunen, J. K. (1990) *J. Biol. Chem. 265*, 2446.

Palosaari, P. M., Vihinen, M., Mäntsälä, P. I., Alexson, S. E. H., Pihlajaniemi, T., & Hiltunen, J. K. (1991) *J. Biol. Chem. 266*, 10750.

Pearson, W. R., & Lipman, D. J. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 2444.

Piatigorsky, J., O'Brien, W. E., Norman, B. L., Kalumuck, K., Wistow, G. J., Borras, T., Nickerson, J. M., & Wawrousek, E. F. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 3479.

Reichert, J., Sakaitani, M., & Walsh, C. T. (1992) *Biochemistry* (in press).

Rhodes, W. C., & McElroy, W. D. (1958) *J. Biol. Chem. 233*, 1528.

Rusnak, F., Faraci, W. S., & Walsh, C. T. (1989) *Biochemistry 28*, 6827.,

Rusnak, F., Sakaitani, M., Drueckhammer, D., Reichert, J., & Walsh, C. T. (1991) *Biochemistry 30*, 2916.

Sacchettini, J. C., Frazier, M. W., Chiara, D. C., Banaszak, L. J., & Grant, G. A. (1988) *Biochem. Biophys. Res. Commun. 153*, 435.

Salinas, P. C., Tolmasky, M. E., & Crosa, J. H. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 3529.

Saraste, M. Sibbald, P. R., & Wittinghofer, A. (1990) *Trends Biochem. Sci. 15*, 430.

Savard, P., Pêloquin, L., & Sylvestre, M. (1986) *J. Bacteriol. 168*, 81.

Schneider, B., Müller, R., Frank, R., & Lingens, F. (1991) *J. Bacteriol. 173*, 1530.

Scholten, J. D., Chang, K.-H., Babbitt, P. C., Charest, H., Sylvestre, M., & Dunaway-Mariano, D. (1991) *Science 253*, 182.

Shalom, T. (1989) *EuGene*, Baylor College of Medicine, Houston, TX.

Smith, R. F., & Smith, T. F. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 118.

Staab, J. F., Elkins, M. F., & Earhart, C. F. (1989) *FEMS Microbiol. Lett. 59*, 15.

Suzuki, H., Kawarabayasi, Y., Kondo, J., & Abe, T., Nishikawa, K., Kimura, S., Hashimoto, T., Yamato, T., (1990) *J. Biol. Chem. 265*, 8681.

Tabor, S., & Richardson, C. C. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 1074.

Toh, H. (1990) *Protein Sequences Data Anal. 3*, 517.

van Liempt, H., von Döhren, H., & Kleinkauf, H. (1989) *J. Biol. Chem. 264*, 3680.

Walsh, C. (1979) Enzymatic Reaction Mechanisms, pp 240–242, W. H. Freeman & Co., New York, NY.

Watanabe, K., Yamano, Y., Murata, K., & Kimura, A. (1986) *Nucleic Acids Res. 14*, 4393.

Weckermann, R., Fuerbass, R., & Marahiel, M. A. (1988) *Nucleic Acids Res. 16*, 11841.

Winkler, F. K., D'Arcy, A., & Hunziker, W. (1990) *Nature 343*, 771.

Wood, K. V., Lam, Y. A., Seliger, H. H., & McElroy, W. D. (1989) *Science 244*, 700.

Yang, X.-Y., Schulz, H., Elzinga, M., & Yang, S.-Y. (1991) *Biochemistry 30*, 6788.

Yeh, L. S., Elzanowski, A., Hunt, L. T., & Barker, W. C. (1988) *Comp. Biochem. Physiol. B: Comp. Biochem. 89*, 433.

Zhao, Y., Kung, S. D., & Dube, S. K. (1990) *Nucleic Acids Res. 18*, 6144.