Discovering New Enzymes and Metabolic Pathways: Conversion of Succinate to Propionate by *Escherichia coli*†

Toomas Haller,‡ Thomas Buckel,§ János Rétey,§ and John A. Gerlt*,‡

Department of Biochemistry, University of Illinois, Urbana, Illinois 61801, and Institut für Organische Chemie, Lehrstuhl Biochemie der Universität, Richard Willstätter-Allee, Universität Karlsruhe, D-76128 Karlsruhe, Germany

Received December 16, 1999; Revised Manuscript Received February 8, 2000

ABSTRACT: The *Escherichia coli* genome encodes seven paralogues of the crotonase (enoyl CoA hydratase) superfamily. Four of these have unknown or uncertain functions; their existence was unknown prior to the completion of the *E. coli* genome sequencing project. The gene encoding one of these, YgfG, is located in a four-gene operon that encodes homologues of methylmalonyl CoA mutases (Sbm) and acyl CoA transferases (YgfH) as well as a putative protein kinase (YgfD/ArgK). We have determined that YgfG is methylmalonyl CoA decarboxylase, YgfH is propionyl CoA:succinate CoA transferase, and Sbm is methylmalonyl CoA mutase. These reactions are sufficient to form a metabolic cycle by which *E. coli* can catalyze the decarboxylation of succinate to propionate, although the metabolic context of this cycle is unknown. The identification of YgfG as methylmalonyl CoA decarboxylase expands the range of reactions catalyzed by members of the crotonase superfamily.

The results of genome sequencing projects consistently reveal that approximately 40% of the identified open reading frames (orfs[1]) encode proteins having uncertain or unknown function (at www.tigr.org/tdb/mdb/mdb.html): their sequences are judged not to be related to those of previously characterized proteins. In eubacteria such as *Escherichia coli*, genes can be organized in operons that encode the enzymes of metabolic pathways (*1*). In favorable cases, this organization can provide important clues about the identities of the unknown proteins, especially when the functions of one or more of the proteins encoded by other genes in the operon are known or inferred (*2*). In eukaryotes, the problem of assigning functions to unknown proteins is considerably more difficult because no clues are available from the identities of the proteins encoded by the proximal genes. Identification of the functions of the unknown proteins identified in genome projects constitutes a significant challenge for the postgenomic era of biochemistry.

The existence of mechanistically diverse enzyme superfamilies (*3*, *4*) may provide an alternate solution for the problem of assigning functions to at least some unknown proteins. Members of such superfamilies catalyze different overall reactions that share a common partial reaction or mechanistic attribute, e.g., generation or stabilization of an oxyanion intermediate that otherwise would be too unstable to be kinetically competent. Low levels of sequence identity (between 15 and 40%) may allow the unknown protein to be placed in an enzyme superfamily for which the underlying catalytic strategy is known. This information may provide a sufficient clue to assign the type of reaction catalyzed by the unknown protein, e.g., racemization, dehydration, isomerization, or decarboxylation. However, the identity of the substrate likely cannot be predicted by such analyses.

For example, the members of the enolase superfamily each catalyze a reaction that is initiated by abstraction of the α-proton of a carboxylate anion substrate to generate a stabilized enolate anion intermediate; depending upon the overall reaction that is catalyzed, the intermediate is partitioned to different types of products via racemization/epimerization, dehydration, deamination, or cycloisomerization (*5*). In this superfamily, the functional groups involved in generation, stabilization, and conversion of the enolate anion intermediate to products are located in an $(\beta/\alpha)_8$-barrel domain; the structural features that determine substrate specificity are located in a separate domain comprised of segments from the N- and C-termini. Recognizing this design, the Urbana group has been able to assign functions to several members of the enolase superfamily that were first identified as unknown proteins in genome sequencing project, including D-galactonate (*2*) and L-rhamnonate[2] dehydratases from *Escherichia coli* and L-Ala-D/L-Glu epimerases[3] from both *E. coli* and *Bacillus subtilis*. In these successes, the unknown proteins catalyzed reactions in metabolic pathways that were encoded by operons, thereby providing essential clues for deducing the identities of the substrate.

† This research was supported by the Universität Karlsruhe, the Deutsche Forschungsgemeinschaft, and the Fonds der Chemischen Industrie (to J.R.) and NIH Grants GM-40570 and GM-52594 (to J.A.G.).

* Address correspondence to this author. Telephone: 217-333-3945. Fax: 217-265-0385. E-mail: j-gerlt@uiuc.edu.

‡ University of Illinois.

§ Universität Karlsruhe.

[1] Abbreviations: 4CLD, 4-chlorobenzoyl CoA dehalogenase; Δ3Δ2, $\Delta^{3,2}$-enoyl CoA isomerase; BadI, 2-ketocyclohexylcarboxyl CoA hydrolase; CaiD, carnitinyl CoA epimerase; DIEN, $\Delta^{3,5},\Delta^{2,4}$-dienoyl CoA isomerase; ECHM, rat mitochondrial enoyl CoA hydratase; FHCL, feruloyl CoA hydratase/lyase; FPLC, fast protein liquid chromatography; IPTG, isopropyl β-thiogalactoside, LB, Luria broth; MenB, 1,4-dihydroxynaphthoyl CoA synthase; NMR, nuclear magnetic resonance; orf, open reading frames; PCR, polymerase chain reaction.

[2] B. K. Hubbard, J. Delli, and J. A. Gerlt, unpublished observations.

[3] B. K. Hubbard, D. Z. Schmidt, and J. A. Gerlt, unpublished observations.

10.1021/bi992888d CCC: $19.00
Published on Web 03/29/2000

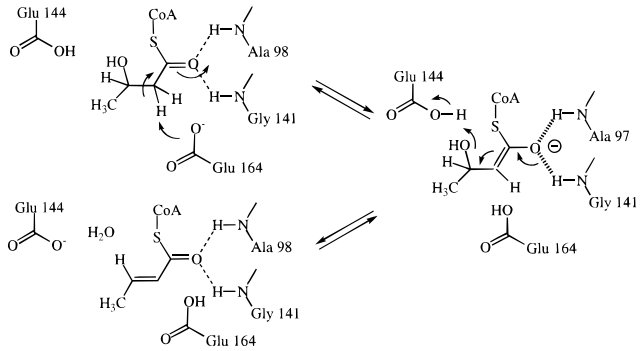


FIGURE 1: Mechanism of the reaction catalyzed by rat mitochondrial crotonase (enoyl CoA hydratase).

The crotonase (enoyl CoA hydratase) superfamily also is mechanistically diverse (*3*, *4*, *6*, *7*). The reaction catalyzed by crotonase involves reversible hydration of enoyl CoAs in β-oxidation. Referring to the rat mitochondrial crotonase that has been structurally characterized (*8*, *9*), two Glu residues function as essential acid/base catalysts (Figure 1): Glu 144 facilitates attack of water on the 3-carbon of the 2-enoyl CoA substrate, and Glu 164 delivers a proton to the 2-carbon to yield the 3-OH acyl CoA product. The thioester carbonyl group of the substrate is bound in an oxyanion hole formed by two peptidic NH groups. Although an enolate anion intermediate stabilized by hydrogen bonding could lie on the reaction coordinate of a stepwise reaction, the available data are consistent with stabilization of an anionic transition state in an asynchronous concerted reaction (*10*, *11*).

Even though the mechanisms of most of the diverse reactions catalyzed by members of the crotonase superfamily have not been investigated, their identities suggest that each involves an oxyanion that must be stabilized to be kinetically competent (*3*, *4*, *6*, *7*). Enolate anion intermediates, or transition states that resemble enolate anion intermediates, are expected to occur in the reactions (Figure 2) catalyzed by crotonase (vide infra; ECHM), $\Delta^{3,2}$-enoyl ($\Delta3\Delta2$) and $\Delta^{3,5},\Delta^{2,4}$-dienoyl CoA (DIEN) isomerases, carnitinyl CoA epimerase (CaiD), 1,4-dihydroxynaphthoyl CoA synthase (MenB), 2-ketocyclohexylcarboxyl CoA hydrolase (BadI), and feruloyl CoA hydratase/lyase (FCHL). An anionic Meisenheimer intermediate (a mimic of an enolate anion intermediate) occurs in the reaction catalyzed by 4-chlorobenzoyl CoA dehalogenase (4CLD), and an anionic tetrahedral intermediate likely occurs in the reaction catalyzed by 3-hydroxyisobutyryl CoA hydrolase (HIBCH), although a mechanism involving an enolate anion intermediate cannot be discarded without experimentation. High-resolution X-ray structures are available for rat mitochondrial crotonase (ECHM; *8*, *9*), 4CLD (*12*), and DIEN (*13*). Each reveals the presence of an oxyanion hole formed by two peptidic NH groups (*3*, *4*, *6*, *7*); these occur in consensus sequences (Figure 3), suggesting that an oxyanion hole is crucial to the reaction catalyzed by each member of the crotonase superfamily. These hydrogen bond donors can be expected to provide significant stabilization of the oxyanion intermediate as the negative charge is localized on the thioester carbonyl oxygen (*14*−*16*).

In contrast to our level of understanding of structure/function relationships in the enolase superfamily, alignment of the sequences of the members of the crotonase superfamily appears insufficient to identify the acid/base catalysts that

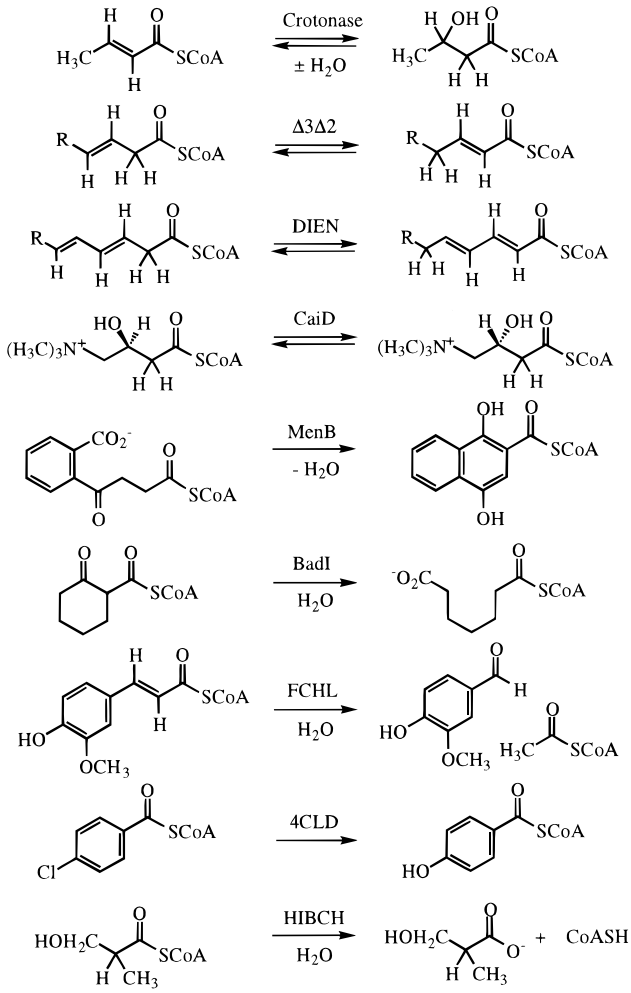


FIGURE 2: Reactions catalyzed by members of the crotonase superfamily.

participate in the formation of the anionic intermediates by unstudied members of the crotonase superfamily. This deficiency likely results from the design of the active site: the secondary structure elements that deliver the acid/base functional groups also provide the specificity determinants for the substrate (*7*−*9*, *12*, *13*). As a result, identification of the chemistries catalyzed by unknown members of the crotonase superfamily is expected to be more difficult than for members of the enolase superfamily.

The *E. coli* genome encodes seven members of the crotonase superfamily (paralogues), four of which have not been characterized. Functional identification and mechanistic characterization of these would provide appropriate tests of whether membership in mechanistically diverse enzyme superfamilies can facilitate identification of unknown proteins.

One of these paralogues, YgfG, is the subject of the studies reported in this and the companion paper. Alignment of the sequence of YgfG with those of other members of the crotonase superfamily suggests the presence of a functional homologue of Glu 144 (Glu 113) but not of Glu 164 in ECHM. This suggests that YgfG catalyzes a new reaction in the crotonase superfamily. The gene encoding YgfG is located in an four-gene operon, *sbm-ygf*D-*ygf*G-*ygf*H. Our in vitro studies suggest that three of the encoded proteins constitute a previously unrecognized adenosylcobalamin-

dependent pathway in *E. coli* in which succinate is decarboxylated to propionate: Sbm is methylmalonyl CoA mutase; YgfG is novel, biotin-independent methylmalonyl CoA decarboxylase; and YgfH is a propionyl CoA:succinate CoA transferase. The product of the fourth gene in the operon, YgfD, likely is a protein kinase that regulates the activity of an as-of-yet unidentified enzyme involved in succinate and/or propionate metabolism.

MATERIALS AND METHODS

All restriction enzymes were purchased from Gibco BRL. T4 DNA ligase and T7 polynucleotide kinase were purchased from New England BioLabs. Oligonucleotide primers were synthesized by Bio-Synthesis (Lewisville, TX), and DNA sequencing was performed by Ana-Gen Technologies Inc. (Palo Alto, CA) and the Genetic Engineering Facility, University of Illinois Biotechnology Center. All other reagents were the highest quality grade commercially available and purchased from Sigma/Aldrich.

Protein mass spectrometry (electrospray) was performed in Mass Spectroscopy Laboratory, School of Chemical Sciences, University of Illinois. ^{1}H NMR spectra were recorded with a Varian Unity INOVA 500 MHz NMR spectrometer. A Bio-Pilot FPLC (Pharmacia Biotech) was used for purification of proteins. Spectrophotometric assays were performed with a Perkin-Elmer Lambda 14 spectrometer. One unit of enzyme activity is the amount that converts 1 μmol of substrate to product per minute.

Cloning, Expression, and Purification of Enzymes. The genes encoding Sbm, YgfD, YgfG, and YgfH were PCR-amplified from chromosomal DNA isolated from *E. coli* strain 1655 using *Pfu* DNA polymerase (Stratagene). Primers containing the first and last 25 bases of each gene were used as primers. For *sbm*, *ygf*D, and *ygf*H, the primers also contained restriction endonuclease cleavage sites at their 5′-ends. The amplification protocol employed 40 cycles of denaturation at 95 °C, annealing at 50 °C, and elongation at 72 °C. The amplified genes were cloned into pET vectors (Novagen) with the following restriction enzymes: *sbm*, pET16-b (5′-*Nde*I and 3′-*Xho*I restriction sites); *ygf*D, pET16-b (5′-*Xho*I and 3′-*Bam*HI restriction sites); *ygf*G, a modification of pET16-b made by replacing the *Nde*I site with a *Stu*I site (cloned with *Stu*I using a blunt-ended strategy); and *ygf*H, a modification of pET15-b in which the N-terminal His-tag contains 10 instead of the usual 6 His residues (5′-*Nde*I and 3′-*Bam*HI restriction sites). All of the resulting plasmids encode proteins with an N-terminal His-tag containing 10 His residues to facilitate purification of the proteins.

All of the proteins were expressed in *E. coli* strain BL21. Transformed cells were grown at 37 °C in Luria broth (LB) to an OD at 600 nm = 2 and harvested by centrifugation (5000 rpm at 4 °C) in a Sorvall JA-10 rotor. No IPTG was added to induce protein expression. The His-tagged proteins were purified with columns (16 cm × 40 cm) of chelating Sepharose Fast Flow (Pharmacia Biotech) charged with Ni^{2+}. Cell lysates were applied to the column in binding buffer (5 mM imidazole, 0.5 M NaCl, and 20 mM Tris-HCl, pH 7.9), washed with wash buffer (60 mM imidazole, 0.5 M NaCl, and 20 mM Tris-HCl, pH 7.9), and eluted with 50% binding buffer/50% strip buffer (100 mM EDTA, 0.5 M NaCl, and 20 mM Tris-HCl, pH 7.9). Using this procedure, Sbm, YgfD, YgfG, and YgfH each were purified to homogeneity based on their electrophoretic behavior on SDS−PAGE.

The molecular weights of the (monomeric) polypeptides of Sbm, YgfD, YgfG, and YgfH were verified by mass spectrometry; the sequences of the encoding genes were confirmed by DNA sequence analysis.

The N-terminal His-tags were removed with Factor Xa (Sbm, YgfD, YgfG) or thrombin (YgfH) according to the supplier's instructions. As isolated, the His-tagged Sbm and YgfD have the sequence GH_{10}SSGHIEGRH appended to their genome-encoded Met N-termini; following cleavage with Factor Xa, the final His of the tag remains so the molecular mass of Sbm is 78 036 kDa and that of YgfD is 37 215 kDa. As isolated, the His-tagged YgfG has the sequence GH_{10}SSGHIEGR appended to its genome-encoded Met N-terminus; following cleavage with Factor Xa, the polypeptide begins with the Met so the molecular mass is 29 173 kDa. Finally, as isolated, the His-tagged YgfH has the sequence GH_{10}GLVPRGSH appended to its genome-encoded Met N-terminus; following cleavage with thrombin, GSH of the tag remains so the molecular mass is 54 106 kDa.

Continuous Spectrophotometric Assay of YgfG (*Methylmalonyl CoA Decarboxylase*). The spectrophotometric assay at 340 nm quantitated the conversion of methylmalonyl CoA to propionyl CoA at 37 °C by monitoring the oxidation of NADH in the presence of oxalacetate, transcarboxylase, and lactate dehydrogenase. With this assay, the methylmalonyl CoA concentration is unchanged as the reaction proceeds. The reactions (1.2 mL) contained YgfG (0.6 μM), methylmalonyl CoA (3−45 μM), 8.3 mM oxalacetate, 0.33 mM NADH, transcarboxylase (5 mU), lactate dehydrogenase (4 mU), and 16.7 mM potassium phosphate, pH 7.2.

HPLC Assay of YgfH (*Propionyl CoA:Succinate CoA Transferase*). The N-terminal His-tag was removed with thrombin after dialyzing YgfH into phosphate-buffered saline. The CoA transferase assays (1 mL) at 25 °C contained propionyl CoA (1−35 μM), 5 mM succinate, and 50 mM sodium HEPES, pH 7. The reactions were initiated by addition of YgfH (1 nM). The reactions were terminated by freezing aliquots in liquid nitrogen. The propionyl CoA that remained and the succinyl CoA that formed were measured by HPLC analysis of reaction aliquots using an reverse-phase C_{18} column (4.6 × 250 mm, Microsorb-MV from Rainin Instruments) at a flow rate of 0.5 mL/min. After injection of the sample and an initial wash with 1% acetonitrile/99% 100 mM ammonium acetate, pH 6.5, for 5 min, the various species were eluted with a linear gradient of acetonitrile and 100 mM ammonium acetate, pH 6.5, in which the acetonitrile concentration was increased from 5% to 25% over 50 min. The retention times of succinyl CoA, methylmalonyl CoA, propionyl CoA, and coenzyme A are 21.6, 22.0, 32.2, and 23.3 min, respectively.

Reconstitution and HPLC Assay of Sbm (Methylmalonyl CoA Mutase). The N-terminal-His-tagged Sbm was reconstituted with adenosylcobalamin using the a modification of the procedure used to reconstitute the methylmalonyl CoA mutase from *Propionibacterium shermanii* expressed in *E. coli* (*17*). In the dark, adenosylcobalamin (50 μM) was added to an ice-cold solution of Sbm (0.3 mg/mL) in 50 mM Tris-HCl, pH 7.5, containing 2 mM dithiothreitol. The pink

FIGURE 3: Partial sequence alignment of the seven members of the crotonase superfamily encoded by the *E. coli* genome and the structurally characterized members of the superfamily.

solution was incubated at 4 °C for 30 min before assays were performed.

The assays (500 μL) were performed at 37 °C and contained succinyl CoA (2−43 μM; *18*−*20*), Sbm (8 nM), and a kinetic excess of YgfG (methylmalonyl CoA decarboxylase; see Results and Discussion) in 50 mM sodium Tris-HCl, pH 7.5, containing 30 mM KCl. The reactions were terminated by freezing aliquots in liquid nitrogen. The succinyl CoA that remained and the propionyl CoA that formed (derived from methylmalonyl CoA by irreversible YgfG-catalyzed decarboxylation) were measured following HPLC of the reaction aliquots using the same protocol described previously for propionyl CoA:succinate CoA transferase (YgfH).

RESULTS AND DISCUSSION

Paralogues of the Crotonase Superfamily Encoded by the E. coli Genome. A BLASTP search (at http://www.ncbi.nlm.nih.gov/blast/blast.cgi) of the proteins encoded by the *E. coli* genome using the sequence of rat mitochondrial crotonase (SwissProt P14604; ECHM) as the query yielded seven paralogues of the crotonase superfamily.

A portion of an alignment (CLUSTAL) of the sequences of the *E. coli* paralogues with those of the structurally characterized rat mitochondrial crotonase (ECHM), 4-chlorobenzoyl CoA dehalogenase (PIR A42560; 4CLD), and rat dienoyl CoA isomerase (SwissProt P23965; DIEN) is presented in Figure 3. This region of sequence includes those secondary structure elements that form the active sites of ECHM, 4CLD, and DIEN. The consensus sequences for the two peptidic NH groups that comprise the oxyanion hole for the thioester carbonyl group/thioester enolate anion are highlighted in green. The catalytically important functional groups in ECHM are Glu 144 and Glu 164 (highlighted in red); those in 4CLD are His 90 (orange) and Asp 145 (blue), and those in DIEN are Asp 176 (red), Glu 196 (red), and Asp 204 (blue).

The functions of three of the paralogues are known: FadB (SwissProt P21177); CaiD, carnitinyl CoA epimerase (SwissProt P31551); and MenB, 1,4-dihydroxynaphthoyl CoA synthase (SwissProt P27290). The remaining four [PaaF (SwissProt P76082), PaaG (SwissProt P77467), YfcX (SwissProt P77399), and YgfG (SwissProt P52045)] have not been characterized, so their functions are unknown or uncertain.

FadB, a multifunctional protein that catalyzes three steps in *E. coli*'s β-oxidation pathway (*21*) contains homologues of Glu 144 and Glu 164 in its crotonase domain. YfcX, a putative multifunctional protein that is 35% identical to FadB, appears to contain a crotonase domain with homologues of both Glu residues in ECHM.

The enantiomers of carnitine as their CoA esters are precursors to crotonobetainyl CoA, an electron acceptor in anaerobic metabolism (CaiD; *22*). The epimerization of carnitinyl CoA catalyzed by CaiD may proceed via a dehydration/rehydration mechanism analogous to that of the crotonase-catalyzed reaction, given homologues for both essential Glu residues in ECHM. However, this pair of functional groups is expected to be insufficient to explain the equilibration of the configuration of the chiral center, e.g., catalysis of both syn and anti dehydration/rehydration reactions presumably is required to accomplish epimerization.

MenB catalyzes the nucleophilic attack of an enolate anion derived from a thioester on a carboxylate group in menaquinone biosynthesis (*23*). The sequence of MenB does not contain homologues of the functional groups found in the structurally characterized members of the crotonase superfamily, so the identities of its active site functional groups are unknown.

The *E. coli* genome encodes a catabolic pathway for phenylacetate (*24*). An operon includes two contiguous genes, *paa*F and *paa*G, that encode members of the crotonase superfamily that are part of this pathway. PaaG likely catalyzes the isomerization of *cis*,*cis*-muconyl CoA to *trans*,*trans*-muconyl CoA, perhaps using a homologue in Asp 204 in DIEN as a nucleophilic catalyst. PaaF contains homologues for both Glu residues in ECHM and likely catalyzes the hydration of *trans*,*trans*-muconyl CoA formed by PaaG to initiate a modified β-oxidation pathway. Neither enzyme has been studied. As detailed in the next section, the seventh paralogue is YgfG.

Genomic Context of YgfG. We have focused our attention on establishing the identity of the reaction catalyzed by YgfG. YgfG contains a homologue of Glu 144 (Glu 113) but not of Glu 164 in ECHM. As noted previously, in ECHM Glu 144 activates water for attack on the 3-carbon of enoyl CoA substrates, and Glu 164 delivers a proton to the 2-carbon of the 3-OH acyl CoA product. To date, no other member of the crotonase superfamily, either of known or unknown

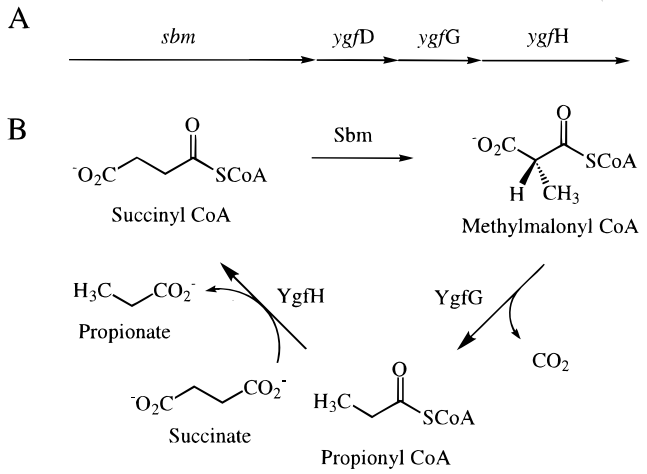


FIGURE 4: (A) Operon containing the gene encoding YgfG. (B) Pathway for the conversion of succinate to propionate in *E. coli*.

function, contains a homologue of Glu 144 but not of Glu 164 in ECHM.

According to the GenBank annotation (GenBank AE000-375), YgfG is encoded by the third gene in a four-gene operon (Figure 4A). The protein encoded by the first gene, Sbm, is 59% identical to the structurally characterized methylmalonyl CoA mutase from *Propionibacterium shermanii*. Despite this high level of sequence identity, Sbm has not been thought to catalyze the methylmalonyl CoA mutase reaction (P. F. Leadlay and J. Roth, personal communications): hence, the name Sbm or *S*leeping *b*eauty *m*utase (*25*).

The protein encoded by the second gene, YgfD, contains a consensus binding sequence for ATP and recently was described to catalyze the phosphorylation of two periplasmic binding proteins involved in cationic amino acid transport (*26*). The gene encoding this protein was reported to complement a mutant deficient in transport that was characterized by reduced levels of phosphorylation of the periplasmic binding proteins.

BLASTP searches using the sequence of YgfH, the protein encoded by the fourth gene, as query, and reveals that it is homologous to acyl CoA transferases that utilize acetyl CoA or succinyl CoA as coenzyme A donors and 4-OH butyrate or butyrate as acceptors (*27*).

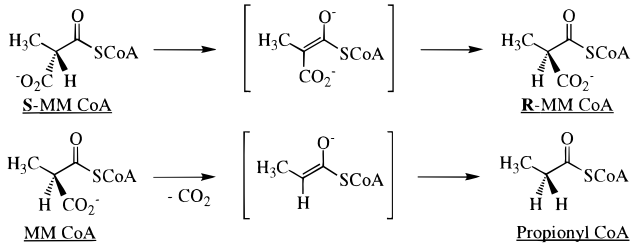
Function of YgfG. The gene encoding YgfG was PCR-amplified and expressed in *E. coli*; YgfG was purified to homogeneity via an N-terminal His-tag as described in the Materials and Methods section. On the basis of the presence of a gene in the same operon that encodes Sbm, a homologue of methylmalonyl CoA mutase, we first assayed YgfG for methylmalonyl CoA epimerase activity. A priori, a member of the crotonase superfamily could be expected to catalyze this reaction because it likely would involve a thioester enolate anion intermediate (*3*, *4*). Methylmalonyl CoA mutases catalyze the isomerization of the (*R*)-diastereomer of methylmalonyl CoA and succinyl CoA (*28*−*31*). In the catabolism of branched chain amino acids and odd-chain fatty acids (*32*), propionyl CoA is carboxylated (by propionyl CoA carboxylase) to the (*S*)-diastereomer of methylmalonyl CoA; the epimerase catalyzes its conversion to the (*R*)-diastereomer so that it can be isomerized to succinyl CoA by methylmalonyl CoA mutase. However, YgfG did not catalyze the methylmalonyl CoA epimerase reaction.

Table 1: Kinetic Constants

protein	k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m (M^{-1} s^{-1})
methylmalonyl CoA decarboxylase (YgfG)	1.6 ± 0.14	13.6 ± 3.3	1.2×10^5
propionyl CoA:succinate CoA transferase (YgfH)[a]	0.72 ± 0.07	7.1 ± 1.9	1×10^5
methylmalonyl CoA mutase (Sbm)[b]	0.2 ± 0.02	11.2 ± 1.6	1.8×10^4

[a] Propionyl CoA varied; succinate constant at 5 mM. [b] Succinyl CoA substrate.

Instead, YgfG was found to inhibit reactions that utilize methylmalonyl CoA as substrate, including transcarboxylase that is used to assay methylmalonyl CoA epimerase. Given the fact that such inhibition could be produced by decarboxylation of methylmalonyl CoA, we assayed YgfG for this activity. The decarboxylation of methylmalonyl CoA, like the epimerization of methylmalonyl CoA, would involve the formation of a thioester enolate anion intermediate that would be stabilized in the oxyanion hole deduced to be present in this member of the crotonase superfamily.



As explained below, we later determined that methylmalonyl CoA epimerase is not encoded by the *E. coli* genome.

As quantitated in Table 1, YgfG catalyzes the efficient decarboxylation of methylmalonyl CoA, with the value of k_{cat}/K_m in the range expected for an evolved protein catalyst (*33*). No attempt was made to optimize the conditions of this assay, so the values we report for both k_{cat} and k_{cat}/K_m can be considered estimates of their lower limits.

The previously characterized methylmalonyl CoA decarboxylases occur in membrane-associated multiprotein complexes that utilize the free energy derived from decarboxylation to transport sodium (*34*−*36*). These complexes include a biotin-containing protein as well as enzymes that catalyze the transfer of the carboxyl group to and from the biotin cofactor. YgfG has no detectable sequence similarity with these proteins and also lacks the consensus sequence for covalent acylation by biotin (Prosite sequence 00188).

Although YgfG catalyzes the efficient decarboxylation of methylmalonyl CoA, we were concerned that this could be an adventitious/promiscuous activity of the protein we isolated: the decarboxylation of methylmalonyl CoA should be a facile reaction in an active site that is designed to stabilize an enolate anion derived from a CoA thioester. This concern was accentuated by our recent discovery that the *o*-succinylbenzoate synthase from *Amycolaptosis* also catalyzes the inefficient racemization of N-acylamino acids, a reaction that we believe is metabolically unimportant (*37*). As such, we also purified and studied the possible reactions catalyzed by YgfH, Sbm, and YgfD to establish whether

these would facilitate a description of a metabolic context for the decarboxylation of methylmalonyl CoA.

Function of YgfH. The gene encoding YgfH was PCR-amplified and expressed in *E. coli*; YgfH was purified to homogeneity via an N-terminal His-tag as described in the Materials and Methods section. The His-tagged YgfH lacked catalytic activity. After cleavage of the His-tag, the "native" YgfH was catalytically active but had limited solubility.

We hypothesized that (i) YgfH catalyzes a CoA transferase reaction and (ii) its substrate specificity would provide clues both as to the physiological context of the decarboxylation reaction catalyzed by YgfG as well as the identities of the substrates and/or products of the reactions catalyzed by both Sbm and YgfD. Recall that Sbm was thought not to catalyze the methylmalonyl CoA mutase reaction.

We observed that YgfH catalyzes a CoA transferase reaction: succinyl CoA and propionyl CoA are the preferred coenzyme A donors; succinate and propionate are the preferred acceptors, with both acetate and butyrate serving as less efficient acceptors. No attempt was made to optimize the conditions of our assay. Despite these caveats, the value we determined for k_{cat}/K_m (Table 1) using propionyl CoA as donor and succinate as acceptor is consistent with those reported for homologous CoA transferases. We conclude that YgfH is propionyl CoA:succinate CoA transferase.

Function of Sbm. On the basis of the ability of YgfG to catalyze the decarboxylation of methylmalonyl CoA to propionyl CoA and of YgfH to catalyze the transfer of coenzyme A from propionyl CoA to succinate, we suspected that Sbm *must* be methylmalonyl CoA mutase, despite "lore" to the contrary. The gene encoding Sbm was PCR-amplified and expressed in *E. coli*; Sbm was purified to homogeneity via an N-terminal His-tag as described in the Materials and Methods section. When the His-tagged Sbm was reconstituted with adenosylcobalamin according to the procedure described for reconstitution of the methylmalonyl CoA mutase from *P. shermanii* expressed in *E. coli* (*17*), we observed both the conversion of methylmalonyl CoA to succinyl CoA and the conversion of succinyl CoA to methylmalonyl CoA using an HPLC assay. [*E. coli* does not synthesize adenosylcobalamin de novo, although it can synthesize the cofactor from exogenous hydroxocobalamin (*38*). Thus, the Sbm purified from *E. coli* is an apoenzyme that must be reconstituted with exogenous adenosylcobalamin.] We subsequently determined that reproducible reconstitution of apoSbm requires only adenosylcobalamin and dithiothreitol.

Although the equilibrium position of the isomerization reaction lies in the direction of succinyl CoA [the reported equilibrium constant is 23.5 (*39*, *40*)], our HPLC protocol does not cleanly separate methylmalonyl CoA and succinyl CoA, making quantitation of the values of kinetic constants by this procedure difficult. However, in the presence of YgfG (methylmalonyl CoA decarboxylase), the quantitative conversion of succinyl CoA to propionyl CoA via methylmalonyl CoA, but without its accumulation, could also be demonstrated; propionyl CoA is well-separated from both succinyl CoA and coenzyme A with our HPLC protocol. Thus, we determined the values of the kinetic constants for the Sbm-catalyzed isomerization by quantitating the formation of propionyl CoA in this coupled-enzyme assay.

The kinetic constants for the Sbm-catalyzed reaction for the conversion of succinyl CoA to methylmalonyl CoA are included in Table 1. Using the reported equilibrium constant for the interconversion of methylmalonyl and succinyl CoAs (23.5), we calculate that the value of k_{cat}/K_m for the conversion of methylmalonyl CoA to succinyl CoA is 4.2×10^5 M^{-1} s^{-1}. This latter value is comparable to the value of 5.7×10^5 M^{-1} s^{-1} reported recently (*41*) for the conversion of methylmalonyl CoA to succinyl CoA by the methylmalonyl CoA mutase from *P. shermanii*. We therefore conclude that Sbm *is* methylmalonyl CoA mutase.

Sbm has not been reported to catalyze the methylmalonyl CoA mutase reaction. Perhaps the activity we measured can be attributed to the presence of the N-terminal His-tag we used to isolate the protein: after cleavage of the His-tag, the level of mutase activity that we measure is <7% that observed with the His-tag present. Other prokaryotic methylmalonyl CoA mutases are dimeric proteins, so the N-terminal His-tag we use for purification may favor an oligomerization that is necessary for catalytic activity. We have not investigated further why attempts by others to measure Sbm's methylmalonyl CoA mutase activity have failed.

Identification of a Methylmalonyl CoA Epimerase. With the assistance of David Graham, Department of Microbiology, University of Illinois, we have observed that YqjC, a protein of unknown function encoded by the *Bacillus subtilis* genome, catalyzes the methylmalonyl CoA epimerase reaction (data not shown). The gene encoding YqjC is proximal to the gene encoding that is a homologue of the transcarboxylase components of acetyl/propionyl CoA carboxylases and biotin-dependent methylmalonyl CoA decarboxylases, a location that is consistent with a carboxylation reaction involving the required epimerization of (*S*)- and (*R*)-methylmalonyl CoAs, although the metabolic context of the epimerization reaction is unknown. Homologues of YqjC are encoded by genes adjacent to those encoding biotin-dependent methylmalonyl CoA decarboxylases in *Archaeglobus fulgidus* and *Veillonella parvula*. YqjC has no close homologues in *E. coli*, although it is a member of the vicinal oxygen chelate superfamily (*3*, *4*) that includes glyoxalase I, the only member of this superfamily encoded by the *E. coli* genome. Our observations are consistent with a report that methylmalonyl CoA epimerase from *P. shermanii* is optimally activated by Co^{2+} and is a dimer of identical 16.5 kDa polypeptides (*42*).

Cycle for the Conversion of Succinate to Propionate. On the basis of the activities we have demonstrated for YgfG, YgfH, and Sbm, we predict that the operon that contains their genes likely encodes a cycle for the decarboxylation of succinate to propionate (Figure 4B). That the *E. coli* genome apparently does not encode either propionyl CoA carboxylase (as evidenced by the lack of a homologue for both prokaryotic and eukaryotic carboxylases) or methylmalonyl CoA epimerase (vide infra) requires that this methylmalonyl CoA mutase and the pathway in which it functions have metabolic roles that differ from those involved in valine and odd-chain fatty acid catabolism.

We were unable to measure any effect of hydroxocobalamin on the growth rate and cell yield of *E. coli* under several conditions, including (i) aerobic growth with either succinate or propionate as carbon source and (ii) anaerobic growth with

succinate as carbon source in the presence of either fumarate or trimethylamine *N*-oxide as electron acceptor. We also were unable to detect any effect of hydroxocobalamin on the identities and amounts of fermentation products when *E. coli* was grown anaerobically with glucose as carbon source and the content of the culture medium was determined by ^{1}H NMR spectroscopy [succinate, lactate, acetate, and formate but not propionate are produced in the same amounts both in the absence and presence of hydroxocobalamin (*43*)]. Hence, the metabolic context of the decarboxylation of succinate in *E. coli* is unknown.

Function of YgfD. The gene encoding YgfD was PCR-amplified and expressed in *E. coli*; YgfD was purified to homogeneity via an N-terminal His-tag as described in the Materials and Methods section. Although YgfD has been assigned the ArgK function in the regulation of cationic amino acid transport (*26*), we view this assignment with suspicion given that the gene encoding YgfD is located within an operon that encodes enzymes that convert succinate to propionate (vide infra; Figure 4A).

The reactions catalyzed by methylmalonyl CoA mutase (former Sbm), methylmalonyl CoA decarboxylase (former YgfG), and propionyl CoA:succinate CoA transferase (former YgfH) provide a mechanism for the continued utilization of succinate once catalytic amounts of succinyl or propionyl CoA are available. As a result, we hypothesized that a source of succinyl CoA other than that provided by α-ketoglutarate dehydrogenase complex and/or succinate thiokinase might be required for significant succinate decarboxylation. Because YgfD has a consensus ATP binding sequence, we reasoned that it might be a succinate (or propionate) CoA ligase. However, we were unable to detect the formation of succinyl CoA/propionyl CoA, AMP, or ADP from succinate/propionate, ATP, and coenzyme A in the presence of YgfD.

We also hypothesized that YgfD and YgfG function together as a novel (biotin-independent) propionyl CoA carboxylase: for example, YgfD could form carboxyphosphate from ATP and bicarbonate, and YgfG could generate an enolate anion from propionyl CoA that would be carboxylated. YgfD did not catalyze a measurable ATPase reaction either in the absence or in the presence of YgfG, bicarbonate, and/or propionyl CoA. YgfD and YgfG also were unable to catalyze the formation of methylmalonyl CoA from propionyl CoA, bicarbonate, and ATP. Hence, we conclude that YgfD and YgfG do not function together to carboxylate propionyl CoA.

A remaining hypothesis is that YgfD encodes a protein kinase/phosphatase that is involved in regulating the activities of other enzymes in *E. coli* which utilize succinyl CoA, propionyl CoA, succinate, and/or propionate as substrates. As precedent for this type of regulation, the flux of carbon through the glyoxalate cycle in *E. coli* is reversibly regulated by the action of isocitrate dehydrogenase kinase/phosphatase with phosphorylation causing inhibition and, as a result, an increase in the concentration of isocitrate (*44*).

CONCLUSIONS

We have demonstrated that a previously unknown and uncharacterized operon in the *E. coli* genome encodes enzymes that decarboxylate succinate to propionate, although the metabolic function of this pathway remains uncertain. YgfG, a previously unidentified member of the crotonase superfamily, catalyzes the decarboxylation of methylmalonyl CoA to propionyl CoA. Sbm catalyzes the methylmalonyl CoA mutase reaction. YgfH catalyzes the interconversion of propionyl CoA and succinate, providing persuasive evidence that these are the substrates and products of the pathway encoded by the operon.

The methylmalonyl CoA decarboxylation reaction catalyzed by YgfG represents an expansion of the chemistries known to be catalyzed by members of the crotonase superfamily. Structures of YgfG determined both in the absence of ligands and in the presence of an inert analogue of methylmalonyl CoA are reported in the companion paper (*45*).

ACKNOWLEDGMENT

We thank Dr. Patricia C. Babbitt, University of California, San Francisco, for valuable discussions and Ms. Ingeborg Merkler for technical assistance in Karlsruhe.

REFERENCES

1. Blattner, F. R., Plunkett, G., Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goeden, M. A., Rose, D. J., Mau, B., and Shao, Y. (1997) *Science 277*, 1453−1474.
2. Babbitt, P. C., Mrachko, G. T., Hasson, M. S., Huisman, G. W., Kolter, R., Ringe, D., Petsko, G. A., Kenyon, G. L., and Gerlt, J. A. (1995) *Science 267*, 1159−1161.
3. Babbitt, P. C., and Gerlt, J. A. (1997) *J. Biol. Chem. 272*, 30591−30594.
4. Gerlt, J. A., and Babbitt, P. C. (1998) *Curr. Opin. Chem. Biol. 2*, 607−612.
5. Babbitt, P. C., Hasson, M., Wedekind, J. E., Palmer, D. R. J., Barrett, W. C., Reed, G. H., Rayment, I., Ringe, D., Kenyon, G. L., and Gerlt, J. A. (1996) *Biochemistry 35*, 16489−16501.
6. Wu., W. J, Andersen, V. E., Raleigh, D. P., and Tonge, P. J. (1997) *Biochemistry 36*, 2211−2220.
7. Xiang, H., Luo, L., Taylor, K. L., and Dunaway-Mariano, D. (1999) *Biochemistry 38*, 7638−7653.
8. Engel, C. K., Mathieu, M., Zeelen, J. P., Hiltunen, J. K., and Wierenga, R. K. (1996) *EMBO J. 15*, 5135−5145.
9. Engel, C. K., Kiema, T. R., Hiltunen, J. K., and Wierenga, R. K. (1998) *J. Mol. Biol. 275*, 847−859.
10. Bahnson, B. J., and Anderson, V. E. (1989) *Biochemistry 28*, 4173−4181.
11. Bahnson, B. J., and Anderson, V. E. (1991) *Biochemistry 30*, 5894−5906.
12. Benning, M. M., Taylor, K. L., Liu, R.-Q., Yang, G., Xiang, H., Wesenberg, G., Dunaway-Mariano, D., and Holden, H. M. (1996) *Biochemistry 35*, 8103−8109.
13. Modis, Y., Rilppula, S. A., Novikov, D. K., Norledge, B., Hiltunen, J. K., and Wierenga, R. K. (1998) *Structure 6*, 957−970.
14. Gerlt, J. A., and Gassman, P. G. (1993) *J. Am. Chem. Soc. 115*, 11552−11570.
15. Gerlt, J. A., and Gassman, P. G. (1993) *Biochemistry 32*, 11943−11952.
16. Shan, S. O., and Herschlag, D. (1996) *Proc. Natl. Acad. Sci. U.S.A.* 93, 14474−14479.
17. McKie, N., Keep, N. H., Patchett, M. L., and Leadlay, P. F. (1990) *Biochem. J. 269*, 293−298.
18. Simon, E. J., and Shemin, D. (1953) *J. Am. Chem. Soc. 75*, 2520.
19. Padmakumar, R., Gantla, S., and Banerjee, R. (1993) *Anal. Biochem. 214*, 318−320.
20. Padmakumar, R., Padmakumar, R., and Banerjee, R. (1997) *Methods Enzymol. 279*, 220−224.

21. Clark, D. P., and Cronan, J. E. (1996) In *Escherichia coli and Salmonella: Cellular and Molecular Biology* (Neidhart, F. C., Curtis, R., Ingraham, J. L., Lin, E. C. C., Low, K. B., Magasanik, B., Reznikoff, W. S., Riley, M., Schaechter, M., and Umbarger, H. E., Eds.) Chapter 21, ASM Press, Washington, DC.
22. Eichler, K., Bourgis, F., Buchet, A., Kleber, H. P., and Mandrand-Berthelot, M. A. (1994) *Mol. Microbiol. 13*, 775−786.
23. Meganathan, R. (1996) In *Escherichia coli and Salmonella: Cellular and Molecular Biology* (Neidhart, F. C., Curtis, R., Ingraham, J. L., Lin, E. C. C., Low, K. B., Magasanik, B., Reznikoff, W. S., Riley, M., Schaechter, M., and Umbarger, H. E., Eds.) Chapter 39, ASM Press, Washington, DC.
24. Ferrandez, A., Minambres, B., Garcia, B., Olivera, E. R., Luengo, J. M., Garcia, J. L., and Diaz, E. (1998) *J. Biol. Chem. 273*, 25974−25986.
25. Roy, I., and Leadlay, P. F. (1992) *J. Bacteriol. 174*, 5763−5764.
26. Celis, R. T., Leadlay, P. F., Roy, I., and Hansen, A. (1998) *J. Bacteriol. 180*, 4828−4833.
27. Sohling, B., and Gottschalk, G. (1996) *J. Bacteriol. 178*, 871−880.
28. Sprecher, M., Clark, M. J., and Sprinson, D. B. (1964) *Biochem. Biophys. Res. Commun. 15*, 581−587.
29. Retey, J., and Lynen, F. (1964) *Biochem. Biophys. Res. Commun. 16*, 358−361.
30. Retey, J., and Lynen, F. (1965) *Biochem. Z. 342*, 256−271.
31. Arigoni, D., Lynen, F., and Retey, J. (1966) *Helv. Chim. Acta 49*, 311−316.
32. Voet, D., and Voet, J. G. (1995) *Biochemistry*, 2nd ed., John Wiley & Sons, Inc., New York.
33. Albery, W. J., and Knowles, J. R. (1976) *Biochemistry 15*, 5631−5640.
34. Hilpert, W., and Dimroth, P. (1991) *Eur. J. Biochem. 195*, 79−86.
35. Huder, J. B., and Dimroth, P. (1993) *J. Biol. Chem. 268*, 24564−24571.
36. Bott, M., Pfister, K., Burda, P., Kalbermatter, Ol., Woehlke, G., and Dimroth, P. (1997) *Eur. J. Biochem. 250*, 590−599.
37. Palmer, D. R. J., Garrett, J. B., Sharma, V., Meganathan, R., Babbitt, P. C., and Gerlt, J. A. (1999) *Biochemistry 38*, 4252−4258.
38. O'Toole, G. A., Rondon, M. R., Trzebiatowski, J. R., Suh, S.-J., and Escalante-Semerena, J. C. (1996) In *Escherichia coli and Salmonella: Cellular and Molecular Biology* (Neidhart, F. C., Curtis, R., Ingraham, J. L., Lin, E. C. C., Low, K. B., Magasanik, B., Reznikoff, W. S., Riley, M., Schaechter, M., and Umbarger, H. E., Eds.) Chapter 47, ASM Press, Washington, DC.
39. Kellermeyer, R. W., Allen, S. H. G., Stjernholm, R., and Wood, H. G. (1964) *J. Biol. Chem. 239*, 2562−2569.
40. Kellermeyer, R. W., and Wood, H. G. (1969) *Methods Enzymol. 13*, 207−215.
41. Padmakumar, R., and Banerjee, R. (1995) *J. Biol. Chem. 270*, 9295−9300.
42. Leadlay, P. F. (1981) *Biochem. J. 197*, 413−419.
43. Alam, K. Y., and Clark, D. P. (1989) *J. Bacteriol. 171*, 6213−6217.
44. LaPorte, D. C., and Cronan, J. E. (1996) In *Escherichia coli and Salmonella: Cellular and Molecular Biology* (Neidhart, F. C., Curtis, R., Ingraham, J. L., Lin, E. C. C., Low, K. B., Magasanik, B., Reznikoff, W. S., Riley, M., Schaechter, M., and Umbarger, H. E., Eds.) Chapter 16, ASM Press, Washington, DC.
45. Benning, M. M., Haller, T., Gerlt, J. A., and Holden, H. M. (2000) *Biochemistry 39*, 4630−4639.

BI992888D