1990

Primary Sequence of the *Escherichia coli fadBA* Operon, Encoding the Fatty Acid-Oxidizing Multienzyme Complex, Indicates a High Degree of Homology to Eucaryotic Enzymes

Concetta C. DiRusso
Center for Metabolic Disease, Ordway Research Institute, cdirusso2@unl.edu

JOURNAL OF BACTERIOLOGY, Nov. 1990, p. 6459–6468
0021-9193/90/116459-10$02.00/0

Vol. 172, No. 11

Primary Sequence of the *Escherichia coli fadBA* Operon, Encoding the Fatty Acid-Oxidizing Multienzyme Complex, Indicates a High Degree of Homology to Eucaryotic Enzymes

CONCETTA C. DiRUSSO

Department of Biochemistry, University of Tennessee, Memphis, Tennessee 38163

Received 4 June 1990/Accepted 28 August 1990

In *Escherichia coli* at least five enzyme activities required for the beta-oxidation of fatty acids are associated with a multienzyme complex composed of two subunits in $\alpha_2\beta_2$ conformation (A. Pramanik et al., J. Bacteriol. 137:469–473, 1979). In the present work, the DNA sequence of the genes encoding these two subunits, *fadB* and *fadA*, has been determined. The direction of transcription was from *fadB* to *fadA* rather than from *fadA* to *fadB*, as suggested previously (S. K. Spratt et al., J. Bacteriol. 158:535–542, 1984). Only 10 nucleotides separated the coding sequences for the two peptides, confirming the suggestion that these genes form an operon. The peptides encoded by *fadB* and *fadA* were 729 amino acids and 387 amino acids, respectively, in length. The larger and smaller peptides had predicted molecular masses of 79,678 and 40,876 Da, respectively. Recently, the sequence of the *fadA* gene was published in a separate report (Yang et al., J. Biol. Chem. 265:10424–10429, 1990). In this work, most of the DNA sequence for *fadA* was confirmed, and 10 errors were corrected. Three of these nucleotide changes resulted in five amino acid residue changes predicted in the carboxy terminus of the *fadA*-encoded peptide. By comparison to other peptide sequences, the α subunit encoded within *fadB* had 31% perfect identity with the rat peroxisomal enoyl-coenzyme A:hydratase-3-hydroxyacyl-coenzyme A dehydrogenase trifunctional enzyme over the entire length of the two peptides. In agreement with the work of Yang et al., the β subunit encoded within *fadA* had 35 to 45% perfect identity with five thiolase genes from different eucaryotic sources over the entire length of the peptide.

Inducible fatty acid-oxidizing systems are characteristic of procaryotes and eucaryotic microsomal systems. Growth of hepatocytes or yeasts in culture in medium containing long-chain fatty acids causes a proliferation of peroxisomes and an induction of the enzymes required for long-chain fatty acid oxidation (29, 31). A similar response is seen after analysis of tissues isolated from animals fed a high-fat diet (19, 20). Current evidence is accumulating that the induction of fatty acid-oxidizing enzyme activities noted in peroxisomes is a result of increased mRNA levels (13). The best-characterized system of inducible fatty acid oxidation is that described for *Escherichia coli*, in which many structural genes and a regulatory gene, *fadR*, required for the beta-oxidation of long- and medium-chain fatty acids have been characterized by classical and molecular genetic techniques (9, 15, 21, 23, 25, 28). Parallels between the peroxisomal and *E. coli* systems can be drawn both at the level of regulation by induction and in terms of enzyme structure. In each case, fatty acyl-coenzyme A (CoA) thiolase activity is associated with a single polypeptide, while at least three additional enzyme activities required for the beta-oxidation of fatty acids are associated with a single multifunctional polypeptide.

In *E. coli*, a multifunctional enzyme complex encoded within the *fadBA* operon exhibits enoyl-CoA hydratase (EC 4.2.1.17), 3-hydroxyacyl-CoA dehydrogenase (EC 1.1.1.35), *cis*-Δ^3-*trans*-Δ^2-enoyl-CoA isomerase (EC 5.3.3.3), 3-hydroxyacyl-CoA epimerase, and 3-ketoacyl-CoA thiolase (EC 2.3.1.16) activities. The complex is composed of two subunits in an $\alpha_2\beta_2$ conformation. The two identical large subunits (α) are approximately 78,000 Da, encoded within the *fadB* gene, and the two identical small subunits (β) are approximately 42,000 Da, encoded within the *fadA* gene (25). Thiolase activity is specifically associated with the small subunit, while the remaining four activities appear to be associated with the large subunit (25). The enzymes of the multienzyme complex have broad substrate specificities for long- and medium-chain fatty acid derivatives except for the enoyl-CoA hydratase, which appears to be specific for medium-chain substrates (25).

In this report, the nucleotide sequence of a 4.1-kb DNA fragment encoding the *fadBA* operon is presented. This work was initially undertaken to identify the *fadBA* promoter which is negatively controlled by the product of the *fadR* gene. The position of the genes within the cloned DNA confirmed previous work (28), but the direction of transcription was shown to be from *fadB* to *fadA* rather than from *fadA* to *fadB*. Recently, the sequence of the *fadA* gene was also published by Yang et al. (35). The sequence presented here mostly substantiates the amino acid sequence of the thiolase gene predicted by Yang et al. (35) with the exception of five amino acid residues in the carboxy terminus, which I found to have been incorrectly reported.

I have found a striking degree of homology between the protein products of the *E. coli* genes *fadB* and *fadA* and several eucaryotic counterparts, suggesting that the structure and function of these proteins are highly conserved. The *E. coli* system offers a powerful molecular approach by which to dissect the enzyme activities associated with each of these peptides.

MATERIALS AND METHODS

Strains and plasmids. *E. coli* JM103 [Δ(*lac pro*) *thi strA endA sbcB hsdR*(F′ *traD36 proAB lacI*^{q}ZΔM15) (17)] was used for the propagation of M13 derivatives. C600 (F$^-$ *thi-1 leuB6 lacY1 tonA21 supE44* λ$^-$) was used for the propagation of the plasmids and to generate strains SR1, SR2, and SR3, described below. Liquid cultures were grown at 37°C in 2YT

(16 g of Bacto tryptone [Difco], 8 g of yeast extract, and 5 g of sodium chloride per liter). When minimal medium was required, M9 (18) medium, usually containing 25 mM dextrose, was used. When necessary, ampicillin was added to 100 μg/ml. BL21(DE3)(plysS) was used for the expression of the *fadBA* operon transcribed by T7 RNA polymerase essentially as described by Studier and Moffat (30).

Enzymes and radiochemicals. Restriction enzymes and Sequenase were purchased from U.S. Biochemicals. [α-^{35}S]dATP and [^{35}S]methionine were purchased from New England Nuclear Corp.

Identification of plasmid-encoded proteins. For the identification of proteins and to confirm the direction of transcription, the T7 system of Studier and Moffat (30) was used. A 4.0-kb *Hin*dIII fragment encoding the *E. coli fadR* gene was inserted into the vector pT7-5 (obtained from S. Tabor, Department of Biological Chemistry, Harvard University, Boston, Mass.) so that *fadR* was transcribed in the opposite direction from the T7 RNA polymerase-responsive promoter. The resultant plasmid was termed pCD130. pCD130 was digested with *Bam*HI, and the 9.4-kb *Bam*HI DNA fragment encoding *fadBA* was inserted. The presence of *fadR* stabilized the multicopy plasmids that also encoded *fadBA*. The two possible orientations of the insert were identified by restriction mapping. One was termed pCD140 and the other was called pCD141. These plasmids are illustrated in Fig. 2A and B. The plasmids were transformed into strain BL21(DE3)(plysS). T7 polymerase was induced by IPTG (isopropylthio-β-D-galactopyranoside), and proteins expressed under T7 polymerase control were identified by [^{35}S]methionine incorporation (30). Proteins in cell extracts (usually from 50 μl of cell culture) were displayed on a 12% polyacrylamide–sodium dodecyl sulfate (SDS) gel, and [^{35}S]methionine-labeled proteins were detected by autoradiography.

Construction and enzymatic analyses of mutants carrying insertions in the *fadB* coding region. The plasmid pCD141 was digested with *Bal*I restriction endonuclease, which cleaves the plasmid once within the *fadB* coding region. A 1.3-kb DNA fragment encoding a gene conferring kanamycin resistance (Kan^r cassette; purchased from Pharmacia LKB) was ligated into the *Bal*I site. Two plasmids were isolated, pCD143.1, which carries the Kan^r cassette in the same direction of transcription as the T7 promoter, and pCD143.4, which carries the Kan^r cassette in the opposite direction of transcription from the T7 promoter. These plasmids were tested for the expression of the *fadBA*-encoded peptides and the Kan^r peptide by using the T7 expression system described above.

To generate *E. coli* strains carrying the insertion mutations on the chromosome, the plasmids were linearized and used to transform a *recB recC sbc* strain as described by Winans et al. (33). The mutations were then moved into strain C600 by phage P1 transduction. The C600 derivative of pCD143.1 was termed SR1, and the C600 derivative of pCD143.4 was termed SR2. A third strain, SR3, was generated by inserting the Kan^r cassette in the *Sal*I site of pCD140. In this construction, the *fadBA* genes are uninterrupted but a linked gene, ORF3, is insertionally inactivated. Location of the Kan^r cassette in SR1, SR2 and SR3 was confirmed by P1 transduction. For these experiments, strain DC529 (5) was stabilized with λPL209. The stabilized fusion strain, λΦ(*fadA-lacZ*), was transduced with P1 *vir* grown on SR1 (Kan^r::*fadB1*), SR2 (Kan^r::*fadB2*), or SR3(*zif*::Kan^r). Transductants were selected on LB plates (18) containing kanamycin (20 μg/ml) and subsequently tested for the ability to grow on minimal medium containing lactose, an indicator of the expression of Φ(*fadA-lacZ*), or inability to grow on minimal medium containing oleate, an indicator of a mutation in *fadB*.

To assay the enzyme activities encoded within the *fadBA* genes, C600, SR1, and SR2 were grown in 100 ml of TB (18) containing 5 mM oleate and 0.5% Brij 58 to mid-log phase and rinsed once with 25 ml of 0.2 M KPO_4 (pH 8) containing 0.5% Brij 58 and once with 25 ml of 0.2 M KPO_4 (pH 8) buffer alone. The washed cells were resuspended in 3 ml of the same buffer and lysed by passage through a French pressure cell at 12,000 lb/in^2. The lysate was centrifuged at 40,000 rpm for 15 min in the TLS55 rotor of a Beckman TL-100 centrifuge. Glycerol was added to 25% and dithiothreitol was added to 2 mM. A sample was removed, heat treated for 1 min at 70°C, and assayed immediately for β-hydroxyacyl-CoA-dehydrogenase activity with 0.1 mM NADH–30 μM acetoacetyl-CoA as the substrate (3). The remaining cell extract was frozen at −20°C and assayed on the following day for thiolase activity with 20 μM acetoacetyl-CoA as the substrate and for crotonase activity with 30 μM crotonyl-CoA as the substrate (3). Protein concentrations were estimated by the procedure of Lowry et al. (16) with bovine serum albumin as the standard. Results are the average of three experiments in which samples were assayed in triplicate.

Dideoxy sequencing. The 9.4-kb *Bam*HI fragment encoding *fadBA* (28) was digested with *Bgl*II, and the fragments were cloned into *Bam*HI-digested M13mp19. Clones were identified for each fragment in either of the two possible orientations except for one 876-bp *Bgl*II fragment, which was only obtained in one orientation. For this fragment, the opposite strand was sequenced by using the plasmid pCEM and in vitro-synthesized oligonucleotides as primers. Nested deletions were prepared for clones with inserts greater than 500 bp in length as described previously (9). For two *Bgl*II clones of 513 bp in length, an *Acc*I deletion from the *Acc*I site of M13mp19 to an *Acc*I site within the clone was constructed to facilitate sequencing beyond 300 nucleotides. The sequence across all junctures in both directions was confirmed by plasmid sequencing of pCEM with *fadBA*-specific oligonucleotides as primers. All primers were synthesized on a Pharmacia LKB Gene Assembler Plus. All reagents for DNA synthesis were purchased from Pharmacia. For dideoxy sequencing, reagents were purchased as a Sequenase V.2 kit from U.S. Biochemicals. Regions of sequence ambiguities due to premature termination or compressions were resolved by the addition of Single-Stranded Binding Protein (purchased from U.S. Biochemicals) during the labeling reaction or by using reaction mixtures containing dITP in place of dGTP.

Computer analysis. All sequence data were collated and analyzed with the programs of the Genetics Users Group of the University of Wisconsin, Madison (8).

Nucleotide sequence accession number. The *fadBA* sequence has been assigned GenBank accession no. M36149.

RESULTS

DNA sequence of the *fadBA* operon. Previous work had positioned the coding region for the *fadBA* operon within a 9.4-kb *Bam*HI fragment of *E. coli* DNA (28). In this work, I have sequenced this DNA fragment and identified the coding regions for two polypeptides of 729 amino acids and 387 amino acids, corresponding to the *fadB* and *fadA* genes, respectively. The complete DNA sequence and predicted

amino acid sequence of the two peptides are presented in Fig. 1. The predicted molecular masses of these peptides, 79,678 and 40,876 Da, were in reasonable agreement with estimates generated by SDS-polyacrylamide gel electrophoresis of the isolated protein complex (2). The predicted amino-terminal sequence for the two peptides corresponded exactly to the results of automated Edman degradation analysis published by Yang et al. (35): MLYKGDTLYL for the α subunit encoded within *fadB* and MEQVVIVDAI for the β subunit encoded within *fadA*. The positions of the two open reading frames and the molecular masses corresponded quite closely to those predicted by subcloning and complementation studies of the *fadA* and *fadB* gene products as defined by Spratt et al. (28). This sequence information substantiates the nucleotide sequence of the thiolase gene published by Yang et al. (35), with the exception of 10 nucleotides, as presented in Fig. 1. There was a cytosine rather than a thymidine residue at position 2636; additional nucleotides were reported by Yang et al. after nucleotides 3634, 3637, 3648, 3691, 3701, 3706, 3708, and 3713; and a cytosine residue was missing after nucleotide 3823. The discrepancies at positions 3634, 3637, and 3648 change a stretch of five amino acid residues. Yang et al. published the amino acid sequence GLADGCVSGL; I propose that the correct sequence is GLATMCIGL. This change accounts for a 387-residue peptide with a calculated molecular weight of 40,876 rather than a 388-residue peptide with a calculated molecular weight of 40,889, as predicted previously (35).

The coding sequences for the two genes were only 10 nucleotides apart rather than 109 nucleotides apart as predicted by Yang et al. (35). Each was preceded by a series of nucleotides which showed homology to the consensus sequence for the *E. coli* ribosome-binding site (27): AGGAGA for the 79-kDa peptide and AAGGAG for the 41-kDa peptide. The predicted ribosome-binding site for the 41-kDa peptide overlapped the termination codon for the 79-kDa peptide. This termination codon was only 9 nucleotides from the initiation codon for the 41-kDa peptide.

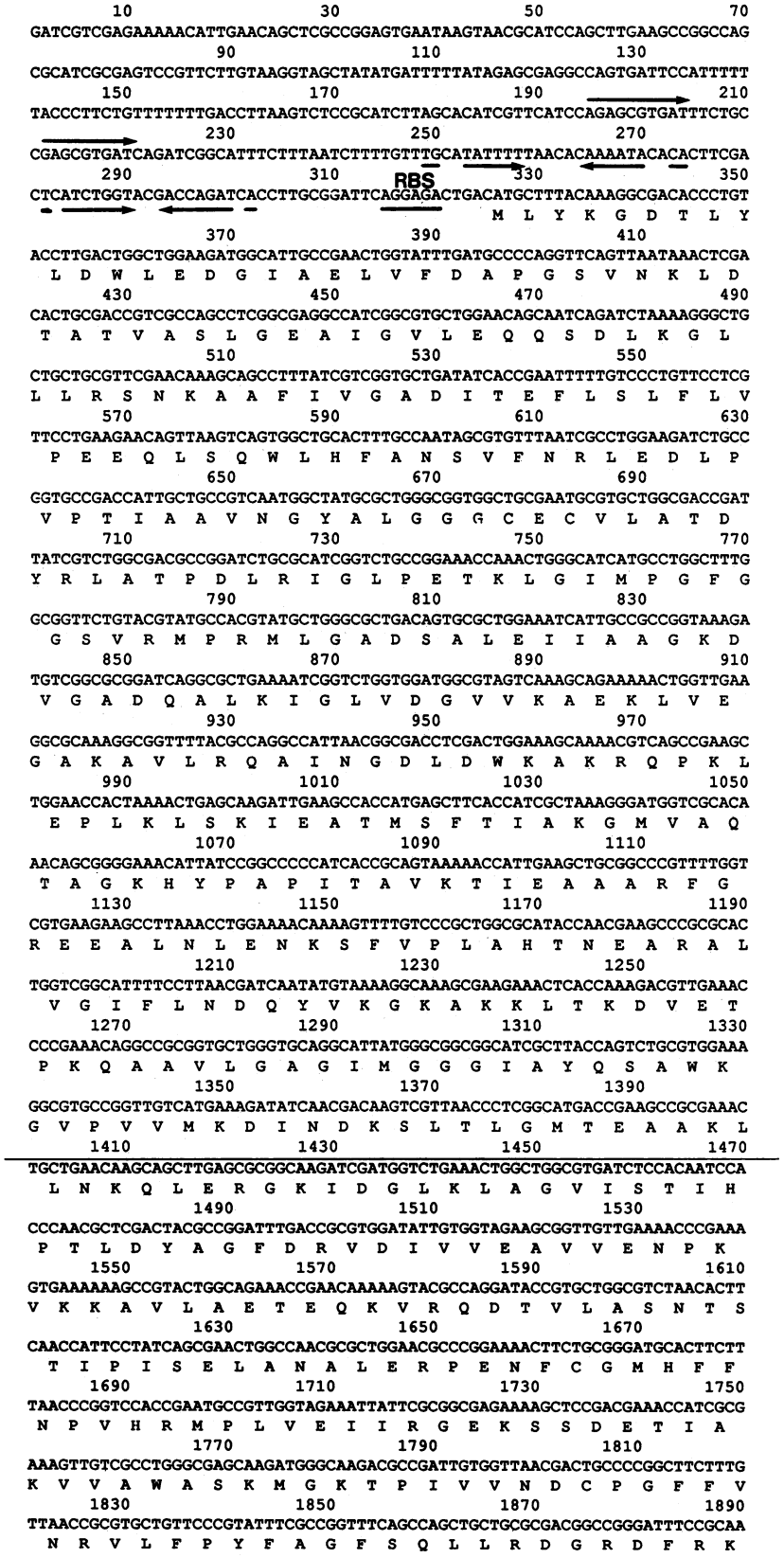
There were several interesting features of the DNA sequence within the predicted promoter region and the predicted region of transcription termination for this operon. The predicted −10 region showed reasonable homology to the consensus sequence for *E. coli* promoters, TATTT, with four of six nucleotides in common. The predicted −35 region showed little homology to the consensus sequence. This was not surprising because the *fadBA* promoter is subject to catabolite repression and therefore requires cyclic AMP (cAMP)-catabolite gene activator protein (CAP) for full activity (5). There were two direct repeats of the sequence 5′-GAGCGTGAT′-3 adjacent to the −35 region which are candidates for cAMP-CAP binding. These sequences showed partial homology to a consensus sequence for cAMP-CAP, 5′-AANTGTGAN$_2$TN$_4$CA (6), where N$_n$ indicates a stretch of any nucleotides. Mutational analysis of this region will be required to define what role, if any, these sequences play in control of *fadBA* expression.

The termination codon for the *fadA* gene was located at nucleotides 3686 to 3689, as illustrated in Fig. 1. A sequence which may be sufficient for factor-independent transcription termination was located 352 nucleotides downstream. This region was G+C rich, with dyad symmetry, and was followed by a series of T residues: CCCGCCTGTCA GGGGCGGGTTTTTTT. The sequence between the termination codon and this terminator structure contained two sets of overlapping direct repeats. The first set was 41 nucleotides long, from nucleotides 3688 to 3729 and 3787 to 3828. The second set was 130 nucleotides long, from nucleotides 3751 to 3881 and 3849 to 3979. Within the same region were eight inverted repeats. The RNA sequence predicted to be encoded within this region of DNA sequence was analyzed by using the computer programs FOLD and SQUIGGLES (8). This analysis predicted that at least one highly stable double-stranded RNA molecule would form (see Fig. 5).

Functional analysis of *fadB* insertion mutants. This sequencing work predicted that the direction of transcription is from *fadB* to *fadA*. Furthermore, identification of the gene products encoded within the plasmid pCD141 after induction of transcription by T7 RNA polymerase clearly showed that the 79- and 41-kDa peptides are transcribed in the same direction as T7 polymerase-catalyzed transcription (see Fig. 3). pCD140 encoded a peptide of approximately 26 kDa in the same direction of transcription as the T7 promoter but in opposite orientation to the 79- and 41-kDa peptides. An open reading frame (termed ORF3) corresponding to this peptide was identified by DNA sequencing beyond the region presented in this work and will be discussed elsewhere.

To verify that the peptides predicted from the DNA sequence analysis encoded the two subunits of the multienzyme complex and possessed the enzyme activities required for beta-oxidation of fatty acids in *E. coli*, a kanamycin resistance cassette was inserted into the predicted coding region for the large subunit in plasmid pCD141. Two plasmids were generated from pCD141, termed pCD143.1 and pCD143.4. These constructs differ only in the orientation of the Kan^r cassette relative to the T7 promoter (Fig. 2A). A third plasmid, pCD144.10, was derived from pCD140 by insertion of the Kan^r cassette in the *SalI* site within the coding sequence for ORF3 (Fig. 2B). Each of these plasmids was analyzed with the T7 expression system. As illustrated in Fig. 3, the 41-kDa peptide and the 79-kDa peptide were absent in pCD143.1 and pCD143.4. However, a new peptide appeared with an intermediate molecular mass. This may be a truncated peptide of the larger subunit, since termination at the coding sequence for the *BalI* site would generate a peptide of 46,350 Da. Similarly, ORF3 was no longer present in pCD144.9 or pCD144.10.

Each of these plasmids was recombined into the *E. coli* chromosome, generating strains SR1 (Kan^r::*fadB1*), SR2 (Kan^r::*fadB2*), and SR3 (*zif*::Kan^r). To verify the location of the Kan^r cassette in the *fadBA* operon, the strain λΦ(*fadA-lacZ*) was transduced to Kan^r by using P1 *vir* grown on SR1, SR2, or SR3. The recipient strain, λΦ(*fadA-lacZ*), was unable to grow on minimal medium containing oleate as the sole carbon and energy source, owing to the *fadA-lacZ* fusion. This phenotype was relieved by transforming the strain with the *fadBA*$^+$ plasmid pCEM (data not shown). Transductants obtained for P1 *vir* grown on SR1 or SR2 were not able to grow on oleate (0 of 118 and 0 of 122, respectively), as expected for insertional inactivation of the *fadBA* operon by either Φ(*fadA-lacZ*) or the Kan^r cassette. Fewer than 2% of the transductants (3 of 118 and 3 of 122 for SR1 and SR2, respectively) retained the ability to grow on minimal medium containing lactose, indicating that the recombination event with Kan^r::*fadB* did not inactivate the *lacZ* fusion. As a control, a similar transductional analysis was performed with P1 *vir* grown on SR3 (*zif*::Kan^r). In this case, at least 87% of the Kan^r transductants (135 of 156) were able to grow on minimal medium containing oleate after 7 days, indicating that the Φ(*fadA-lacZ*) had been replaced by the wild-type *fadBA* genes; 13% retained the ability to grow


RBS

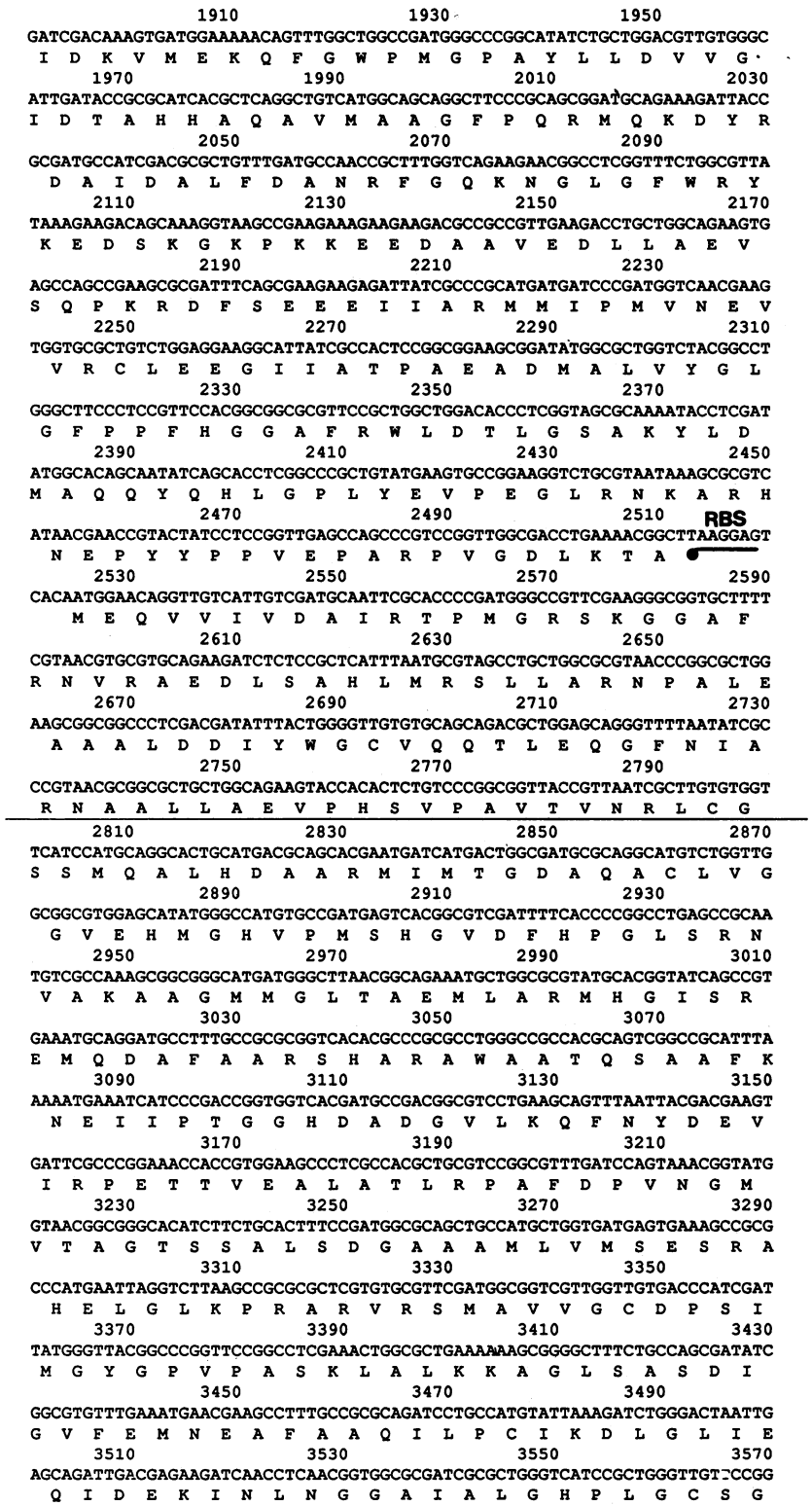


FIG. 1. Nucleotide sequence and derived amino acid sequence of the *fadBA* operon. The codon for the initiating methinine for *fadB* is found at nucleotide 326. The codon for the initiating methinine for *fadA* is found at nucleotide 2525. The ribosome-binding sites (RBS) are underlined. Within the sequence 5′ to the *fadB* gene is a direct repeat, delineated by arrows above the sequence. Two inverted repeats in the region 5′ to the coding region for *fadB* and eight inverted repeats in the region 3′ to the coding sequence for *fadA* are delineated by arrows below the sequence. The region 3′ to the *fadA* gene analyzed with the computer programs SQUIGGLE and FOLD, as illustrated in Fig. 5, is enclosed in a box.

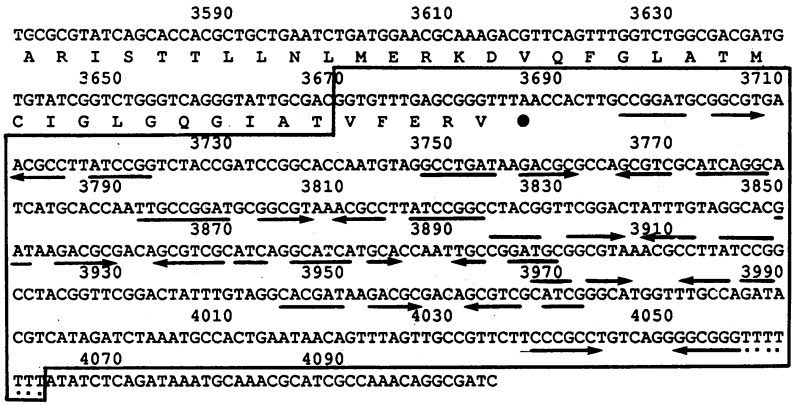


FIG. 1—*Continued*

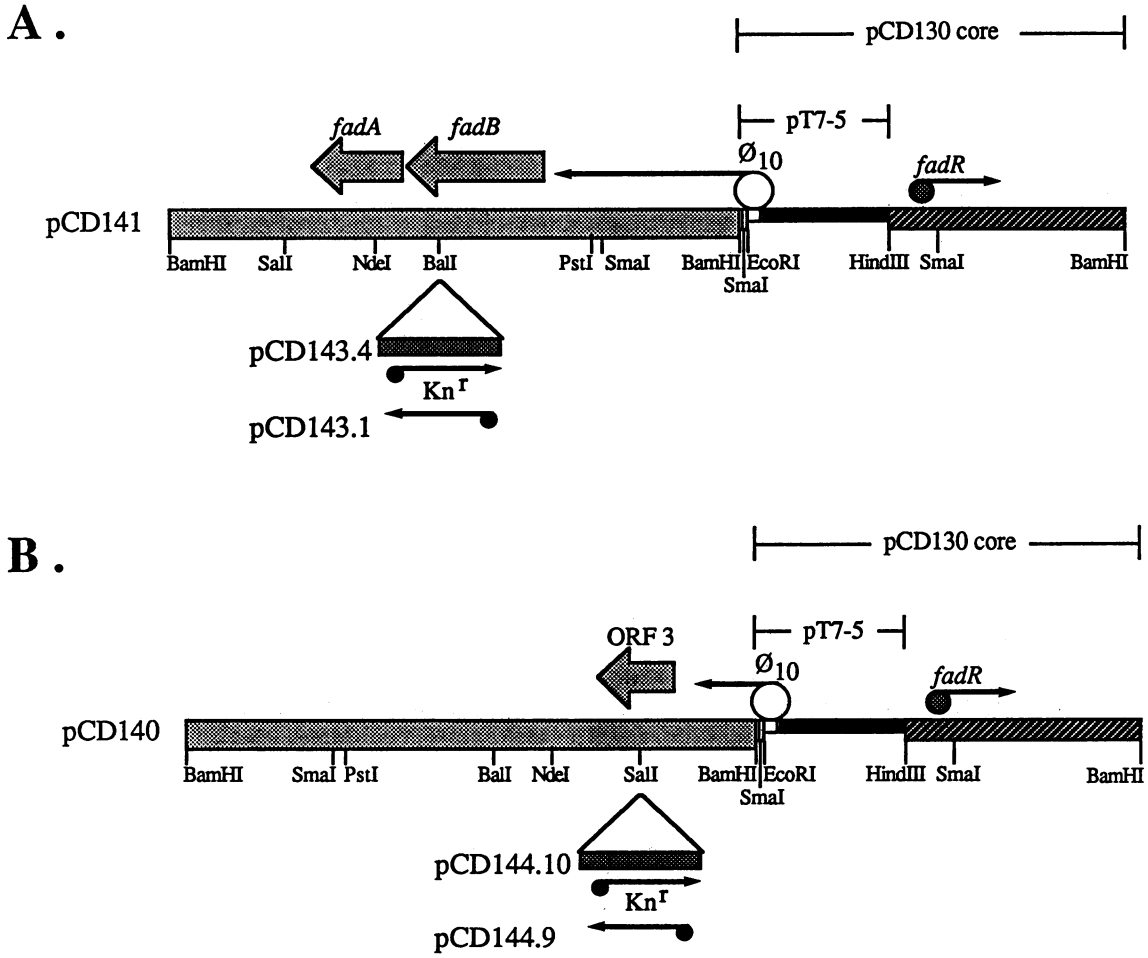


FIG. 2. Restriction maps of plasmids used for the identification and analysis of peptides encoded within the 9.4-kb *fadBA Bam*HI fragment of *E. coli* chromosomal DNA. Construction of each of the plasmids pCD141, pCD140, pCD143.1, pCD143.4, pCD144.9, and pCD144.10 is described in Materials and Methods. The arrow originating with a large open circle indicates the T7 RNA polymerase-transcribed promoter Φ_{10}. The direction of the Kan^r cassette relative to the Φ_{10} promoter is indicated by an arrow originating in a small solid circle. Restriction sites for the 9.4-kb insert correspond to those in pCEM as detailed by Spratt et al. (28). (A) In pCD141, the *fadBA* genes, delineated by large shaded arrows, are transcribed in the same direction as the Φ_{10} promoter. Only the Kan^r cassette (Kn^r) inserted into the *BalI* site of *fadB* is shown for pCD143.1 and pCD143.4. (B) In pCD140, a third open reading frame (ORF3) is indicated by the large shaded arrow. This reading frame is insertionally inactivated in pCD144.9 and pCD144.10, as shown.

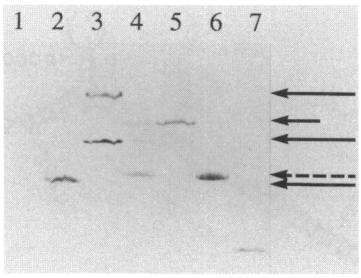


FIG. 3. Identification of proteins encoded within the 9.4-kb *Bam*HI fragment carrying the *fadBA* genes by the T7 expression system of Studier and Moffat (30). Total cell extracts from 50 μl of culture from strain BL21(DE3)(plysS) harboring the plasmid of interest were prepared after IPTG induction and [^{35}S]methionine labeling. The proteins were resolved on a 12% polyacrylamide–SDS gel, and labeled peptides were detected by autoradiography. Lane 1, pCD130; lane 2, pCD140; lane 3, pCD141; lane 4, pCD143.1; lane 5, pCD143.4; lane 6, pCD144.9; and lane 7, pCD144.10. The long arrows denote the three peptides encoded within the *fadBA* insert. The short arrow denotes the presumed truncated *fadB* peptide. The broken arrow denotes the product of the Kan^r cassette.

on lactose (21 of 156), indicating that the Φ(*fadA-lacZ*) had been retained.

To test for insertional inactivation of the *fadBA* gene products, strain SR1, strain SR2, and the parental strain C600 were assayed for three enzyme activities associated with the multienzyme fatty acid-oxidizing complex encoded within *fadBA*. In the parental strain C600, enoyl-CoA hydratase activity, 3-hydroxyacyl-CoA dehydrogenase activity, and 3-ketoacyl-CoA thiolase activity were present at 172, 6.8, and 9.6 nmol/min/mg of protein, respectively. As expected, the *fadB* insertion mutants did not have detectable levels of these enzyme activities.

Protein sequence comparisons with eucaryotic fatty acid-oxidizing proteins. The amino acid sequences derived from the DNA sequence are presented in Fig. 1. These predicted peptide sequences were compared with all sequences in the sequence files of GenBank by using the computer programs TFASTA and BESTFIT. These comparisons identified homologies between the 41-kDa peptide and five thiolase genes from different eucaryotic sources (1, 4, 7, 12, 26) and between the 79-kDa peptide and the gene encoding the rat peroxisomal enoyl-CoA:hydratase-3-hydroxyacyl-CoA dehydrogenase (22). The results of these comparisons are presented graphically in Fig. 4A and B. The 79-kDa subunit of *E. coli* has been shown by Schulz and co-workers to

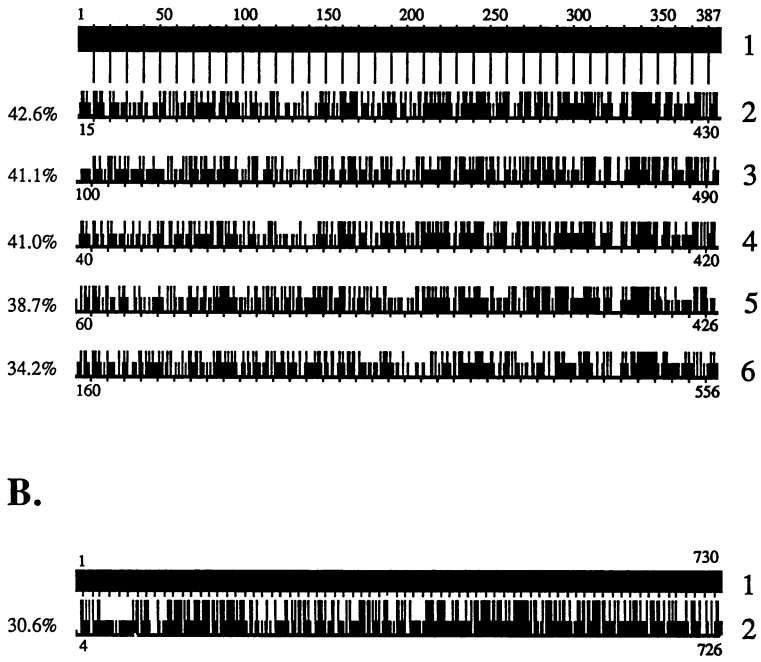


FIG. 4. Amino acid homologies between peptides of the *E. coli* multifunctional fatty acid-oxidizing complex and six eucaryotic peptides. The percent perfect amino acid identity for each comparison of the *E. coli* peptide is listed on the left. Perfect amino acid homologies are indicated by a long bar for a single amino acid; conserved amino acid residue types are indicated by a half bar; nonconserved residue types do not have vertical bars. The position of the first amino acid for each peptide compared with the initiating methionine for *fadA* or *fadB* is listed as a number to the left; the last amino acid included in the comparison is listed to the right. (A) Lane 1, *E. coli fadA* gene encoding β-ketoacyl-CoA thiolase is graphically illustrated as a solid box. It is further divided into 10 amino acid segments as shown. The sources of the eucaryotic thiolase peptides used for comparisons are: lane 2, rat peroxisomal 3-ketoacyl-CoA thiolase (GenBank accession number JO2749 [12]); lane 3, *Zoogloea ramigera* thiolase (GenBank accession number JO2631 [26]); lane 4, human peroxisomal 3-oxoacyl-CoA thiolase (GenBank accession number X12866 [4]); lane 5, rat 3-oxoacyl-CoA thiolase (GenBank accession number XO5341 [1]); and lane 6, *Saccharomyces uvarum* acetoacetyl-CoA thiolase (GenBank accession number XO7976 [7]). (B) A similar comparison is graphically illustrated for (lane 1) the product of the *fadB* gene and (lane 2) the rat peroxisomal enoyl-CoA:hydratase-3-hydroxyacyl-CoA dehydrogenase (GenBank accession number KO3249 [22]).

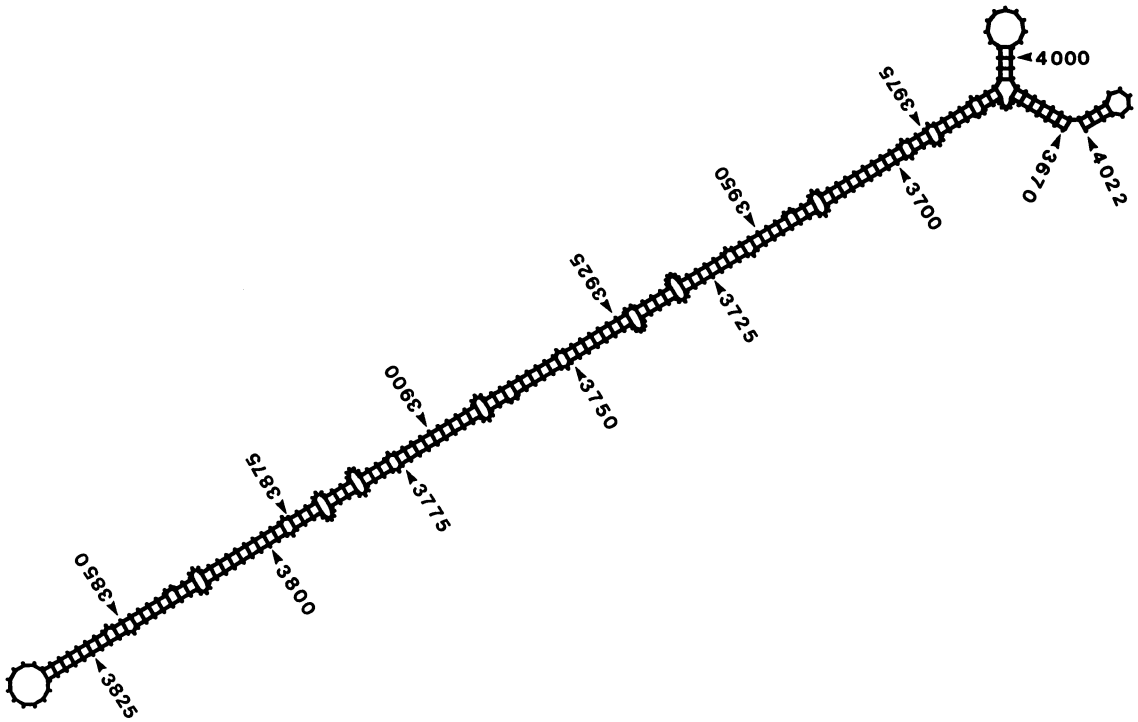


FIG. 5. RNA secondary structure prediction for sequences 3670 to 4022 generated by using the computer programs FOLD and SQUIGGLES (8). This figure indicates one potential RNA structure with a calculated energy of formation of ca. −144.5 kcal.

possess four enzyme activities required for beta-oxidation (2, 25). The amino acid sequence of this peptide had significant amino acid homologies with only one gene in GenBank at this time, the rat peroxisomal enoyl-CoA:hydratase-3-hydroxyacyl-CoA dehydrogenase (22). The rat peroxisomal gene has recently been shown to have a third activity, Δ^3,Δ^2-enoyl-CoA isomerase (24). Therefore, at least three enzyme activities carried out by the *E. coli* multifunctional peptide are also carried out by the peroxisomal enzyme. Comparison of the amino acid sequence for these two peptides showed an overall 31% identity and 54% similarity of residues. The homology occurred over the entire length of the peptides.

DISCUSSION

In this report I present the sequence of the *E. coli fadBA* genes encoding the two subunits of the fatty acid-oxidizing complex. This sequence information confirms and extends the cloning work of Spratt et al. (28) and the nucleotide sequence data for the *fadA* gene presented by Yang et al. (35). There are at least two interesting structural features of these genes. The first is that the coding sequences for the two genes are only 10 nucleotides apart rather than 109 nucleotides apart, as suggested by Yang et al. (35). The termination codon for the *fadB* gene overlaps the predicted ribosome-binding site for the *fadA* gene. This supports the suggestion that these genes form an operon. Additionally, the proximity of the termination codon for *fadB* to the initiation codon for *fadA* may indicate translational coupling, such as that which occurs in the amino acid-biosynthetic operons of *E. coli* (10).

The second interesting structural feature is a series of extensive, nearly perfect, overlapping direct and inverted repeats which are found between the termination codon of the *fadA* gene and a predicted factor-independent transcriptional terminator 352 nucleotides downstream. RNA structural predictions for this region generated by using the computer programs FOLD and SQUIGGLE are presented in Fig. 5. This figure illustrates at least one large stable RNA hairpin. At this time it is difficult to interpret the importance of the direct repeats or predicted RNA hairpin. Perhaps they are important in ensuring transcription termination at the G+C-rich hairpin found between nucleotides 4029 and 4073, as noted above, or in stabilizing the RNA against degradation, or both.

The derived amino acid sequence for the *E. coli* genes encoding the fatty acid-oxidizing multienzyme complex showed a high degree of similarity to the eucaryotic enzymes which have been cloned and sequenced (1, 4, 7, 12, 22, 26). This amino acid homology is expected to reflect an overall similarity of structure and function for these evolutionarily conserved peptides. The *E. coli* system is the best characterized to date both biochemically and genetically. The peptides encoded within the *fadBA* operon form a multienzyme complex, which has been purified (2). Yang et al. have provided kinetic evidence for the channeling of intermediates through the complex (34). The active site for only one of these enzyme activities, that of the thiolase, has been identified (14). This site contains an essential cysteine residue corresponding to amino acid residue 91 of the *E. coli* peptide encoded within *fadA*, as presented here.

Amino acid homologies were found for the large subunit of the *E. coli* multifunctional complex and one peroxisomal protein. The multifunctional peroxisomal protein is similar in structure and function to the *E. coli* peptide encoded within *fadB*, unlike the mitochondrial enzymes, for which each enzyme activity is associated with a unique peptide. Molecular analysis of the genes encoding these proteins, particu-

larly site-directed mutagenesis, will be valuable in evaluating amino acid residues important for enzyme activity and may aid in the dissection of the active site for each enzyme activity. It will also be valuable to contrast these proteins with their mitochondrial counterparts as clones and sequences of these genes also become available.

The *fadBA* operon is inducible by growth in medium containing long-chain fatty acids as the sole carbon and energy source. The expression of these genes is controlled by repression of transcription by the product of the *fadR* gene, FadR. FadR is a multifunctional regulator of fatty acid-related pathways. This regulatory protein negatively controls at least six genes which are required for beta-oxidation of both saturated and unsaturated fatty acids and the *aceBAK* operon, which encodes enzymes of the glyoxylate shunt (reviewed in reference 21). FadR is also a transcriptional activator of at least the *fabA* gene, required for unsaturated fatty acid biosynthesis (11). In previous work, I identified through sequence analysis a helix-turn-helix motif within the *fadR* protein which may be important for DNA binding (9). In the present work, I have identified a region within the *fadBA* promoter with hyphenated dyad symmetry which is a candidate for the *fadR* binding site. I am presently characterizing FadR binding to the *fadBA* promoter in vitro by using purified FadR and DNA fragments. The ultimate goal is to understand how FadR coordinately regulates fatty acid degradation and biosynthesis and acetate metabolism.

ACKNOWLEDGMENTS

The strain BL21(DE3)(plysS) was obtained from William Studier, Brookhaven National Laboratories, Brookhaven, N.Y. The plasmid pT7-5 was obtained from Stan Tabor, Harvard University, Cambridge, Mass. I am grateful for the technical assistance of Shea O. Ross in constructing strains SR1, SR2, and SR3 and for performing the enzyme assays.

This work was supported by Public Health Service grant GM38104 from the National Institutes of Health.

LITERATURE CITED

1. **Arakawa, H., M. Takiguchi, Y. Amaya, S. Nagata, H. Hayashi, and M. Mori.** 1987. cDNA-derived amino acid sequence of rat mitochondrial 3-oxoacyl-CoA thiolase with no transient presequence: structural relationship with peroxisomal isozyme. EMBO J. **6**:1361–1366.
2. **Binstock, J. F., A. Pramanik, and H. Schulz.** 1977. Isolation of a multienzyme complex of fatty acid oxidation from *Escherichia coli*. Proc. Natl. Acad. Sci. USA **74**:492–495.
3. **Binstock, J. F., and H. Schulz.** 1981. Fatty acid oxidation complex from *Escherichia coli*. Methods Enzymol. **71**:403–411.
4. **Bout, A., Y. Teunissen, T. Hashimoto, R. Benne, and J. M. Tager.** 1988. Nucleotide sequence of human peroxisomal 3-oxoacyl-CoA thiolase. Nucleic Acids Res. **16**:10369.
5. **Clark, D.** 1981. Regulation of fatty acid degradation in *Escherichia coli*: analysis by operon fusion. J. Bacteriol. **148**:521–526.
6. **de Crombrugghe, B., S. Busby, and H. Buc.** 1984. Cyclic AMP receptor protein's role in transcription activation. Science **224**: 831–838.
7. **Dequin, S., R. Gloeckler, C. J. Herbert, and F. Boutelet.** 1988. Cloning, sequencing and analysis of the yeast *S. uvarum* ERG 10 gene encoding acetoacetyl-CoA thiolase. Curr. Genet. **13**: 471–478.
8. **Devereux, J., P. Haeberli, and O. Smithies.** 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12**:387–395.
9. **DiRusso, C. C.** 1988. Nucleotide sequence of the *fadR* gene, a multifunctional regulator of fatty acid metabolism in *Escherichia coli*. Nucleic Acids Res. **16**:7995–8009.
10. **Harms, E., E. Higgins, J.-W. Chen, and H. E. Umbarger.** 1988. Translational coupling between the *ilvD* and *ilvA* genes of *Escherichia coli*. J. Bacteriol. **170**:4798–4807.
11. **Henry, M. F., and J. E. Cronan.** 1989. A facile and reversible method to decrease the copy number of the ColE1-related cloning vectors commonly used in *Escherichia coli*. J. Bacteriol. **171**:5254–5261.
12. **Hijikata, M., N. Ishi, H. Kagamiyama, T. Osumi, and T. Hashimoto.** 1987. Structural analysis of cDNA for rat peroxisomal 3-ketoacyl-CoA thiolase. J. Biol. Chem. **262**:8151–8158.
13. **Hijikata, M., J.-K. Wen, T. Osumi, and T. Hashimoto.** 1990. Rat peroxisomal 3-ketoacyl-CoA thiolase gene: occurrence of two closely related but differentially regulated genes. J. Biol. Chem. **265**:4600–4606.
14. **Izbicka-Dimitrijevic, E., and H. F. Gilbert.** 1982. Two sulfhydryl groups near the active site of thiolase I from porcine heart: modification of thiolase with the fluorescent thiol reagent *S*-mercurio-*N*-dansyl-L-cysteine. Biochemistry **21**:6112–6118.
15. **Klein, K., R. Steinberg, B. Fiethan, and P. Overath.** 1971. Fatty acid degradation in *Escherichia coli*: an inducible system for the uptake of fatty acids and further characterization of *old* mutants. Eur. J. Biochem. **19**:442–450.
16. **Lowry, O. H., N. J. Rosebrough, A. L. Farr, and R. J. Randall.** 1951. Protein measurement with the Folin phenol reagent. J. Biol. Chem. **193**:265–275.
17. **Messing, J.** 1983. New M13 vectors for cloning. Methods Enzymol. **101**:20–78.
18. **Miller, J. H.** 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
19. **Neat, C. E., M. S. Thomassen, and H. Osmundsen.** 1981. Effects of high-fat diets on hepatic fatty acid oxidation in the rat. Biochem. J. **187**:149–159.
20. **Neat, C. E., M. S. Thomassen, and H. Osmundsen.** 1980. Induction of peroxisomal β-oxidation in rat liver by high-fat diets. Biochem. J. **186**:369–371.
21. **Nunn, W. D.** 1986. A molecular view of fatty acid catabolism in *Escherichia coli*. Microbiol Rev. **50**:179–192.
22. **Osumi, T., N. Ishii, M. Hijikata, K. Kamaijo, H. Ozosa, S. Furrita, S. Miyazawa, K. Kondo, K. Inoue, H. Kagamiyama, and T. Hashimoto.** 1985. Molecular cloning and nucleotide sequence of the cDNA for rat peroxisomal enoyl-CoA dehydrogenase bifunctional enzyme. J. Biol. Chem. **260**:8905–8910.
23. **Overath, P., G. Pauli, and H. U. Schairer.** 1969. Fatty acid degradation in *Escherichia coli*. An inducible acyl-CoA synthetase, the mapping of *old* mutations, and the isolation of regulatory mutants. Eur. J. Biochem. **7**:559–574.
24. **Palosaari, P. M., and J. K. Hiltunen.** 1990. Peroxisomal bifunctional protein from rat liver is a trifunctional enzyme possessing 2-enoyl-CoA hydratase, 3-hydroxyacyl-CoA dehydrogenase, and Δ^3,Δ^2-enoyl-CoA isomerase activities. J. Biol. Chem. **265**: 2446–2449.
25. **Pawar, S., and H. Schulz.** 1981. The structure of the multienzyme complex of fatty acid oxidation from *Escherichia coli*. J. Biol. Chem. **256**:3894–3899.
26. **Peoples, O. P., S. Masamune, C. T. Walsh, and A. J. Sinskey.** 1987. Biosynthetic thiolase from *Zoogloea ramigera*. III. Isolation and characterization of the structural gene. J. Biol. Chem. **262**:97–102.
27. **Shine, J., and L. Dalgarno.** 1975. Determinant of cistron specificity in bacterial ribosomes. Nature (London) **254**:34–38.
28. **Spratt, S. K., P. N. Black, M. M. Ragozzino, and W. D. Nunn.** 1984. Cloning, mapping, and expression of genes involved in the fatty acid-degradative multienzyme complex of *Escherichia coli*. J. Bacteriol. **158**:535–542.
29. **Spydevold, O., and J. Bremer.** 1989. Induction of peroxisomal β-oxidation in 7800 C1 Morris hepatoma cells in steady state by fatty acids and fatty acid analogues. Biochim. Biophys. Acta **1003**:72–79.
30. **Studier, W. F., and B. A. Moffat.** 1986. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. J. Mol. Biol. **189**:113–130.

31. **Tanaka, A., M. Osumi, and S. Fukui.** 1982. Peroxisomes of alkane-grown yeasts: fundamental and practical aspects. Ann. N.Y. Acad. Sci. **386**:183–199.
32. **Weeks, G., M. Shapiro, R. O. Burns, and S. J. Wakil.** 1969. Control of fatty acid metabolism. I. Induction of the enzymes of fatty acid oxidation in *Escherichia coli*. Eur. J. Biochem. **97**:827–836.
33. **Winans, S. C., S. J. Elledge, J. H. Krueger, and G. C. Walker.** 1985. Site-directed insertion and deletion mutagenesis with cloned fragments in *Escherichia coli*. J. Bacteriol. **161**:1219–1221.
34. **Yang, S.-Y., R. Bittman, and H. Schulz.** 1985. Channeling of a β-oxidation intermediate on the large subunit of the fatty acid oxidation complex from *Escherichia coli*. J. Biol. Chem. **260**: 2862–2868.
35. **Yang, S.-Y., X.-Y. He Yang, G. Healy-Louie, H. Schulz, and M. Elzinga.** 1990. Nucleotide sequence of the *fadA* gene. J. Biol. Chem. **265**:10424–10429.