Nucleotide Sequence of the Promoter and *fadB* Gene of the *fadBA* Operon and Primary Structure of the Multifunctional Fatty Acid Oxidation Protein from *Escherichia coli*[†,‡]

Xue-Ying He Yang,[§] Horst Schulz,*[,§] Marshall Elzinga,[∥] and Song-Yu Yang*[,∥]

Laboratory of Neurobiochemistry, New York State Institute for Basic Research in Developmental Disabilities, Staten Island, New York 10314, and Department of Chemistry, City College of The City University of New York, New York, New York 10031

Received December 26, 1990; Revised Manuscript Received April 12, 1991

ABSTRACT: The primary structure of a multifunctional protein, the large α-subunit of the *Escherichia coli* fatty acid oxidation complex, was determined by sequencing the *fadB* region of the *fadBA* operon. The amino-terminal sequence of this protein had been established by Edman degradation. The transcription start site of the *fadBA* operon was located 42 nucleotides upstream of the initiator codon of the *fadB* gene by primer extension analysis. Sequences of −10 and −35 regions of the promoter responsible for interaction with RNA polymerase were found to be CACACT and TTTGCA, respectively. The location of the promoter of the *fadBA* operon was defined, and the transcription direction of this operon, from *fadB* to *fadA*, as previously proposed [Yang, S.-Y., et al. (1990) *J. Biol. Chem. 265*, 10424–10429], was corroborated. The multifunctional protein is composed of 729 amino acid residues and has a calculated M_r of 79 593. A putative NAD-binding βαβ-fold necessary for L-3-hydroxyacyl-CoA dehydrogenase function was found in the central region of the *fadB* gene product. Sequence analyses suggest that the functional domains of the multifunctional protein are arranged in the order enoyl-CoA hydratase:L-3-hydroxyacyl-CoA dehydrogenase:Δ^3-*cis*-Δ^2-*trans*-enoyl-CoA isomerase and suggest that the genes of the *E. coli* multifunctional protein and rat peroxisomal trifunctional β-oxidation enzyme evolved from a common ancestral gene.

The multienzyme complex of fatty acid oxidation of *Escherichia coli* is encoded by the *fadBA* operon mapped at 87 min of the *E. coli* chromosome (Yang & Schulz, 1983; Yang et al., 1988, 1990; Bachmann, 1990). The nucleotide sequence of the *fadA* gene specifying 3-ketoacyl-CoA thiolase (EC 2.3.1.16), a 42-kDa β-subunit of the multienzyme complex, has recently been reported (Yang et al., 1990). The *fadB* gene product, the α-subunit of the *E. coli* fatty acid oxidation complex, is a 78-kDa multifunctional polypeptide exhibiting four different enzyme activities (Yang & Schulz, 1983; Yang et al., 1988). Enoyl-CoA hydratase (crotonase) (EC 4.2.1.17) and L-3-hydroxyacyl-CoA dehydrogenase (EC 1.1.1.35) catalyze the second and third reactions of the β-oxidation spiral respectively, whereas Δ^3-*cis*-Δ^2-*trans*-enoyl-CoA isomerase (EC 5.3.3.8) and 3-hydroxyacyl-CoA epimerase (EC 5.1.2.3) are auxiliary enzymes that function in the β-oxidation of polyunsaturated fatty acids (Yang et al., 1986a).

The fatty acid oxidation complex overproduced by the expression of the *fadBA* operon within a 5.2-kb (*PstI*–*SalI*) fragment of *E. coli* DNA has been purified and characterized (Yang et al., 1988). The important role of this multienzyme complex in the fatty acid β-oxidation was intensively studied. On the basis of the kinetic results, it was found that β-oxidation intermediates are directly transferred from the active site of enoyl-CoA hydratase (crotonase) to that of L-3-hydroxyacyl-CoA dehydrogenase on the large subunit of the multienzyme complex (Yang et al., 1985, 1986a,b). Because of intermediate channeling, the *E. coli* fatty acid oxidation complex is capable of directly β-oxidizing 2,4-dienoyl-CoAs, key intermediates in the β-oxidation of polyunsaturated fatty acids (Yang et al., 1986a,b). Furthermore, the degradation of 2-*trans*,4-*cis*-decadienoyl-CoA by the *E. coli* fatty acid oxidation complex and by the rat peroxisomal trifunctional β-oxidation enzyme suggested that the epimerase-dependent pathway of unsaturated fatty acid oxidation (Stoffel & Caesar, 1965) is operative in nonmitochondrial β-oxidation systems (Yang et al., 1986a). However, many questions have emerged as a result of these studies. Where on the single polypeptide chain do the four different enzymatic functions of the *fadB* gene product reside? What is the structural basis for the channeling of β-oxidation intermediates on the large subunit of the multienzyme complex? How are the structural genes *fadA* and *fadB* expressed? Information about the *fadBA* promoter and the primary structure of the mutifunctional protein is essential for finding answers to such questions.

In this paper, we report the nucleotide sequence of the 5′ control region of the *fadBA* operon and the primary structure of the *E. coli* multifunctional protein. The transcription start site has been located at an adenine nucleotide 42 bp upstream of the *fadB* initiator codon by primer extension analysis. The sequences of the −10 and −35 regions of the *fadBA* promoter are CACACT and TTTGCA, respectively. The locations of the putative operator and cAMP receptor protein binding sites, relative to the transcription start site, are also reported. Since this work was completed, two reports showing the nucleotide sequence of the *fadA* and *fadB* genes appeared (Nakahigashi & Inokuchi, 1990; DiRusso, 1990). Both confirm our establishment of the transcription direction of the operon (Yang et al., 1990). The data presented here confirm most of the *fadB* sequence of these reports. We extend the work to include an analysis of the enzymatic functional domains within the

[†] This work was supported by the New York State Office of Mental Retardation and Developmental Disabilities; by Grants HL-21471 (to M.E.) and HL-18090 and HL-30847 (to H.S.) from the National Institutes of Health; and by a City University of New York Faculty Research Award.

[‡] The nucleotide sequence reported in Figure 2 of this paper has been submitted to GenBank under Accession Number J05332.

* To whom correspondence should be addressed.

[§] City College of The City University of New York.

[∥] New York State Institute for Basic Research in Developmental Disabilities.

0006-2960/91/0430-6788$02.50/0 © 1991 American Chemical Society

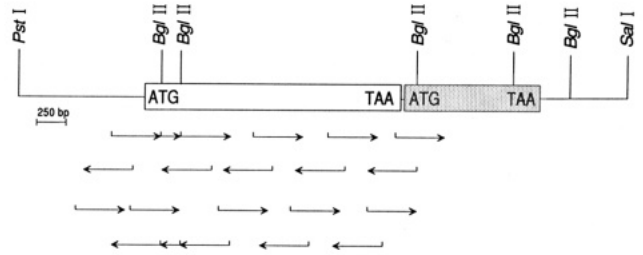

FIGURE 1: DNA sequencing strategy. The coding regions for the multifunctional protein and 3-ketoacyl-CoA thiolase are indicated by the open and shaded boxes, respectively. The restriction sites from which subclones were made for sequencing are shown. The horizontal arrows under the DNA fragments indicate the directions and extents of sequencing.

amino acid sequence of the *E. coli* multifunctional protein. On the basis of sequence homologies, we suggest that exon VII of the rat peroxisomal trifunctional β-oxidation enzyme gene encodes L-3-hydroxyacyl-CoA dehydrogenase and Δ^3-*cis*-Δ^2-*trans*-enoyl-CoA isomerase and that the *E. coli* multifunctional protein and the rat peroxisomal trifunctional β-oxidation enzyme have a common evolutionary origin.

Experimental Procedures

Materials. *E. coli* strain MV1190 is a mutant of JM101 (Yang et al., 1990). *E. coli* strain DH5αF′ competent cells were purchased from Bethesda Research Laboratories. The 5.2-kb DNA fragment containing the *fadBA* operon was purified from *E. coli* strain LS6749 (pK52) after amplification of the recombinant plasmid pK52 as described previously (Yang et al., 1990). M13 mp19 and mp18 have been described by Yanisch-Perron et al. (1985). Restriction endonucleases, T4 polynucleotide kinase, and yeast tRNA were purchased from Boehringer Mannheim. M-MLV reverse transcriptase, RNasin, and RNase-free DNase were from Bethesda Research Laboratories. T4 polymerase and the Cyclone I Biosystem were obtained from International Biotechnologies, Inc. Sequenase Version 2.0 Kit, T4 ligase, and the Klenow fragment of DNA polymerase were purchased from United States Biochemical Corp. Taq DNA polymerase was from Perkin Elmer/Cetus. [α-^{35}S]dATP and [γ-^{32}P]dATP were obtained from Amersham Co. The fluorescent dye labeled primers were supplied by Applied Biosystems, Inc.

DNA Sequencing. The purified 5.2-kb DNA (*Pst*I–*Sal*I) was digested with *Bgl*II. Various restriction fragments were separated by 1% agarose gel electrophoresis, purified from the gel (Volgelstein & Gillespie, 1979), and then cloned into M13 mp19 and mp18 vectors using MV1190 competent cells treated with TSS [Luria-Bertani broth with 10% poly(ethylene glycol) 5% dimethyl sulfoxide, and 50 mM Mg^{2+} at pH 6.5] (Chung et al., 1989). The orientation of an insert in a single *Bam*HI site was ascertained by hybridization assay with single-stranded template DNA (Howarth et al., 1981). Overlapping deletions in single-stranded template DNA were generated with T4 polymerase according to the method of Dale et al. (1985) but using DH5αF′ competent cells as described by Hanahan (1983). The DNA sequencing strategy is shown in Figure 1. The nucleotide sequence was determined by the dideoxy chain-termination method (Sanger et al., 1977; Innis et al., 1988) with 7-deaza-dGTP instead of dGTP (Mizusawa et al., 1986), and base determination was automated by using an Applied Biosystems, Inc., 370A DNA sequencer (Smith et al., 1986). The nucleotide sequence at regions containing compressions was verified manually by use of Sequenase Version 2.0 and [α-^{35}S]dATP, and the substitution of dITP for dGTP (Narnes et al., 1983). Computer-assisted sequence analysis was accomplished by using the PC/GENE (IntelliGenetics Co.) programs (Moore et al., 1988).

Primer Extension Analysis. Total *E. coli* RNA was purified from strain LS6749 (pK52) growing exponentially in an M9 salts medium supplemented with 1% tryptic peptone and 0.1% (v/v) oleate dispersed in 0.4% (v/v) Triton X-100 by a procedure adapted from published methods (Summers, 1970; Gilman & Chamberlin, 1983). A primer specific to *fadB* (5′CGGCAATGCCATCTTCCA3′) was synthesized with an Applied Biosystems, Inc., 380B DNA synthesizer (Caruthers, 1985), and was 5′-end-labeled by using [γ-^{32}P]ATP and T4 polynucleotide kinase as described (Chaconas & van de Sande, 1980). Twenty micrograms of RNA mixed with 400 fmol of ^{32}P-labeled primer (5×10^4 cpm) in 20 μL of buffer (containing 250 mM KCl, 5 mM Tris-HCl, and 0.5 mM EDTA, pH 8.0) was denatured in a heating block set at 80 °C for 2 min, and this sample was then transferred to a 46 °C water bath for hybridization for at least 2 h. After the annealing reaction, the sample was mixed with 20 units of RNasin and was diluted to 125 μL, with final concentrations of 50 mM Tris-HCl (pH 8.3), 3 mM $MgCl_2$, 75 mM KCl, 10 mM DTT, and 2 mM each of dATP, dGTP, dTTP, and dCTP. For each sample 2 μL (400 units) of M-MLV reverse transcriptase was added, and the extension reaction proceeded at 42 °C for 60 min (Williams & Mason, 1985). Two parallel experiments, one with 20 μg of yeast tRNA as a substitute for the RNA samples and another without addition of the reverse transcriptase, were performed as controls. At the end of the primer extension, 4 μL of 0.5 M EDTA was added to stop the reaction. After the nucleic acid was recovered by ethanol precipitation, the pellet was resuspended in 6 μL of formamide gel loading buffer. Then, primer extension products were separated by electrophoresis on an 8% polyacrylamide/8 M urea gel alongside a dideoxy sequencing ladder (Narnes et al., 1983), generated from an M13 mp19 clone of a 1.25-kb *Pst*I–*Bgl*II fragment that includes the initiator codon and 5′-flanking region of the *fadB* gene DNA using the above-mentioned, unlabeled oligonucleotide as primer (see Figure 3).

Results and Discussion

DNA Sequence of the Multifunctional Protein Gene. The nucleotide sequences of both strands at the 5′-flanking region of the *fadA* gene were determined for 2.6 kb (see Figure 1). Inspection of all six reading frames revealed only one with a 2.2-kb open reading frame long enough to accommodate the multifunctional protein gene. The beginning of this open reading frame was located 150 bp upstream of a *Bgl*II site. The complete nucleotide sequence of the *fadB* gene with 381 base pairs of 5′-noncoding region and 12 base pairs in the 3′-noncoding region that links *fadB* to *fadA* is shown in Figure 2. A Shine–Dalgarno sequence is centered nine nucleotides on the 5′-side of the initiator codon of the *fadB* gene. The translation product of the DNA sequence in the 2187 base pair long open reading frame is a polypeptide made up of 729 amino acid residues with a calculated molecular weight of 79 593. This value is in good agreement with the value of 78 kDa of the α-subunit of the *E. coli* fatty acid oxidation complex as determined by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (Pawar & Schulz, 1981). Moreover, we have determined the sequence of the first 10 residues of the α-subunit by Edman degradation, and the result presented in our recent paper (Yang et al., 1990) exactly matches the amino acid sequence of the multifunctional protein deduced from the nucleotide sequence data (see Figure 2). These results indicate that the *fadB* coding region has been precisely identified.

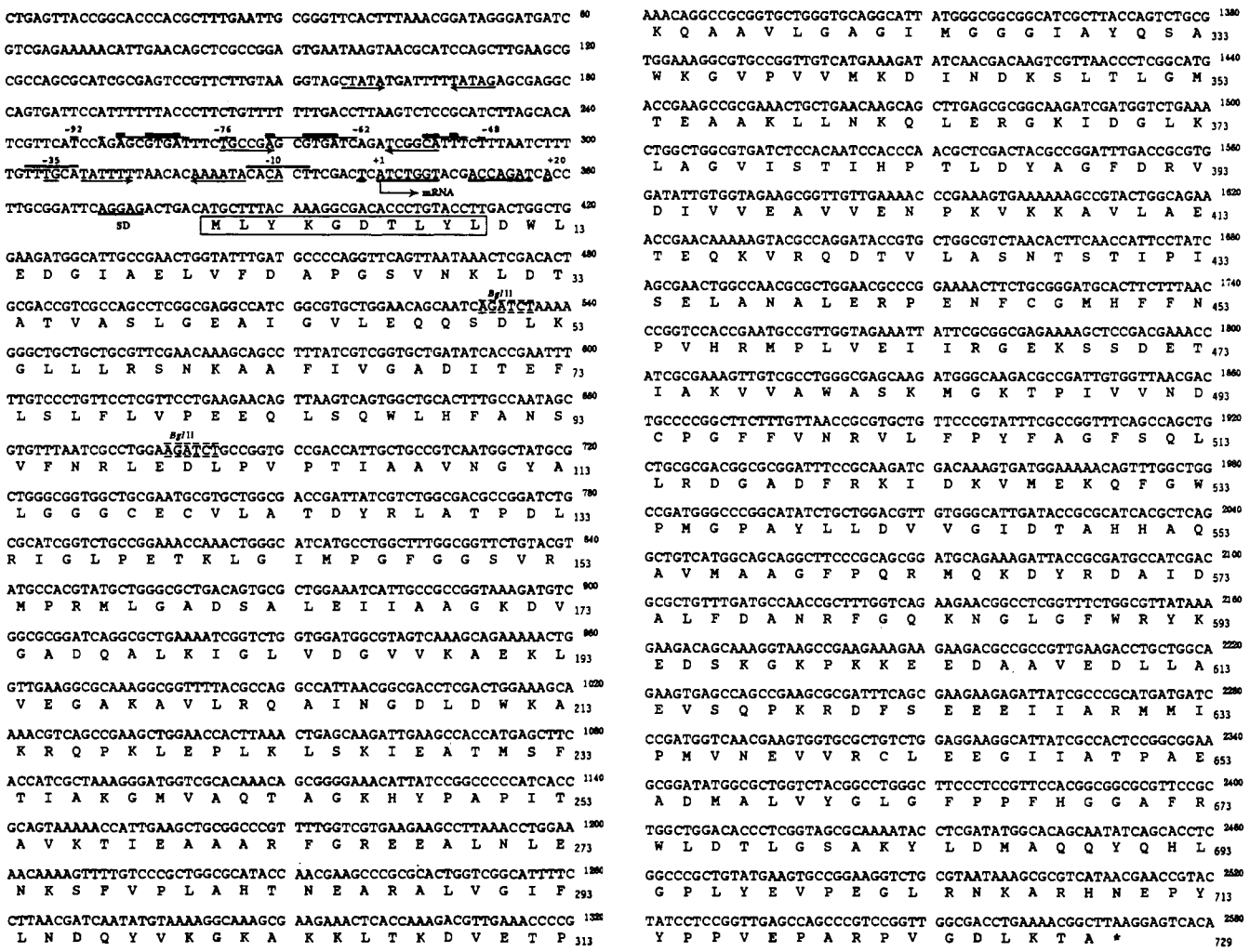

FIGURE 2: Nucleotide sequence of the 5′-noncoding region and *fadB* gene of the *fadBA* operon and the primary structure of the multifunctional protein from *E. coli*. Nucleotides are numbered in the 5′ to 3′ direction. The transcription start site of the *fadBA* operon found at position 340 is marked as +1 above the sequence, and the nucleotides upstream of the transcription start site are given negative numbers for convenience in studying the promoter region. The promoter −10 and −35 regions are indicated by lines above the sequence. Regions showing dyadic symmetry within the putative repressor binding site are underlined. The nucleotide identical with the one found most frequently in the corresponding position of the other known cAMP–CRP binding site (de Crombrugghe et al., 1984) is overlined, and those identical with very highly conserved nucleotides are marked by thick lines. Inverted repeats are delineated by arrows below the sequence. A Shine–Dalgarno sequence is underlined and marked with SD. *Bgl*II restriction sites are indicated by dashed lines. The deduced amino acid sequence (with standard one-letter amino acid abbreviations) is shown below the nucleotide sequence. The amino-terminal 10 amino acids, which are boxed, were determined by Edman degradation (Yang et al., 1990). The stop codon is represented by an asterisk (★).

Thus, the primary structure of the multifunctional protein was deduced (see Figure 2), and the revised organization of the *fad* operon, in which the transcription direction is from *fadB* to *fadA* as first proposed by Yang et al. (1990), has received further support. As shown in Figure 2, we found guanine and cytosine at positions 1933 and 1934, respectively. Our results agree with the data of Nakahigashi and Inokuchi (1990), but conflict with those reported by DiRusso (1990), who found cytosine at position 1933 and guanine at position 1934. We also found cytosines rather than guanines, as reported by Nakahigashi and Inokuchi (1990), at positions 2373 and 2377; at these positions our data agree with the results of DiRusso (1990). In addition, we found guanine and cytosine at positions 120 and 121, respectively, at variance with DiRusso (1990). Four of these nucleotide differences mentioned above occur in the *fadB* coding sequence and result in some differences in the predicted protein sequence: (a) we reported alanine at residue 518, in agreement with the data of Nakahigashi and Inokuchi (1990), rather than arginine as predicted by DiRusso (1990), and (b) we reported the sequence LGFPPFH, here in agreement with DiRusso (1990), rather than LGLPAFH as predicted by Nakahigashi and Inokuchi (1990) at residues 662–668. If the sequences reported in the previous papers are accurate, the observed differences in DNA sequence may reflect genetic heterogeneity among the strains of *E. coli* studied by the different groups.

Identification of the Promoter of the fadBA Operon. The transcription start site was determined by primer extension analysis. The primer was complementary to the *fadBA* mRNA between nucleotides 35 and 52 downstream from the initiator codon of the *fadB* coding sequence (Figure 2). The results of a representative experiment are presented in Figure 3. The size of the most prominent extended fragment indicates that the transcription initiates with an adenine residue (designated +1). The 5′-end of *fadBA* mRNA was clearly mapped at 42 nucleotides upstream from the initiator codon of the *fadB* coding sequence. This result permits identification of the *fadBA* promoter region that interacts with RNA polymerase to form the initiator complex. The −10 region sequence CACACT, with a spacing of 7 bp upstream from the transcription start site, shows a reasonable homology to the consensus Pribnow sequence, and this is very similar to the

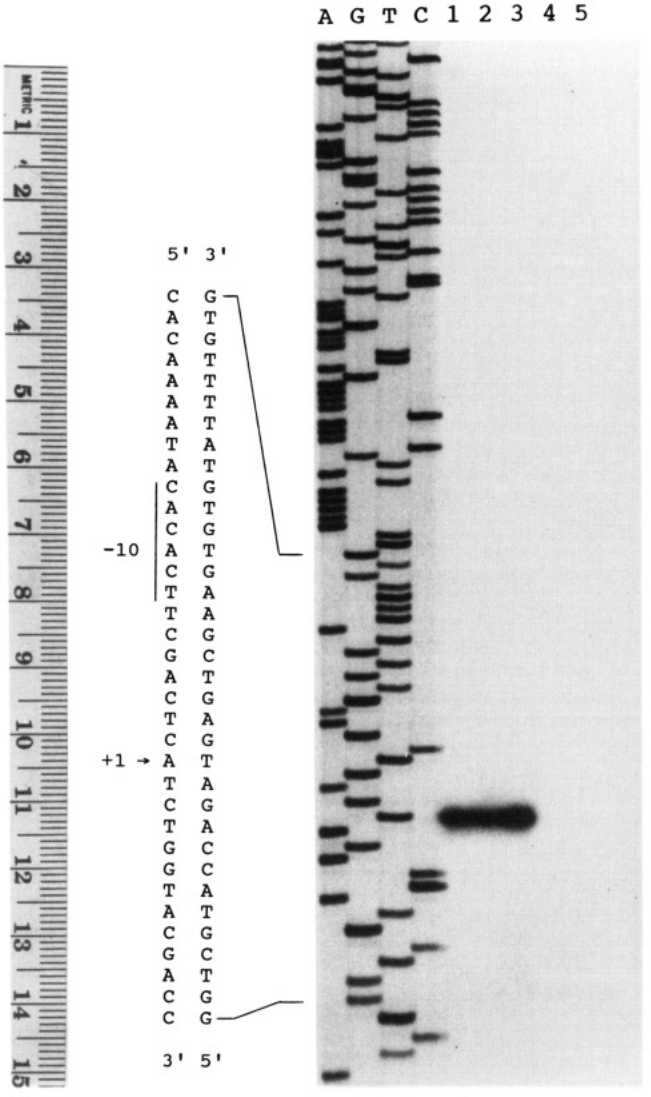

FIGURE 3: Determination of the transcription start site of the *fadBA* operon by primer extension analysis. An oligonucleotide complementary to the *fadB* sequence (from positions 419 to 436, Figure 2) was 5'-end-labeled with ^{32}P, hybridized to RNA isolated from *E. coli* strain LS6749 containing plasmid pK52 after induction by oleate, and extended with reverse transcriptase. For details, see Experimental Procedures. Lanes A, G, T, and C represent the dideoxy sequencing reaction products using the same oligonucleotide as primer and DNA of the *fadB* region as a template. In this experiment, samples were hybridized at 50 °C (lane 1), 46 °C (lanes 2, 4, and 5), and 42 °C (lane 3). Lane 4 is a control run without reverse transcriptase added in the primer extension reaction. Lane 5 is a control in which yeast tRNA was hybridized with the same primer, and the extension reaction was catalyzed by the addition of reverse transcriptase.

−10 region sequence of the *malPQ* operon of *E. coli* (Debarbouille & Railbaud, 1983). Moreover, the −35 region sequence TTTGCA has 4 of 6 nucleotides that are identical with the consensus −35 hexamer. Homology to conserved base pairs around the −35 hexamer (Rosenberg & Court, 1979; Railbaud & Schwartz, 1984) was found from nucleotides −42 to −28 relative to the transcription start site, and the nucleotide sequence upstream of the −35 region is A/T rich. On the basis of the biochemical and genetic evidence, the promoter region of the *fadBA* operon is defined (see Figure 2).

When *E. coli* RNA isolated before oleate induction was used as a template, little primer extension product was detected (data not shown). These results demonstrate that the efficiency of *fadBA* promoter was apparently dramatically raised by the inducer. The finding of the promoter DNA sequence is important for elucidation of the induction mechanism, which is a key to the gene expression of the *fad* regulon (Overath et al., 1969; Nunn, 1986). Sequence analysis of the *fadBA* operon DNA in the region of the transcription start site revealed a 21-bp DNA fragment (from −2 to +19) that shows dyadic symmetry. This region is presumably a repressor binding site, at which the formation of a repressor–operator complex would block the synthesis of *fadBA* mRNA (see Figure 2). In addition, a 26-bp DNA fragment that extends from −10 to −35 exhibits partial twofold symmetry. If a repressor binds to this site, it would interfere with the formation of promoter–RNA polymerase initiation complex (Figure 2). It was reported (Pauli et al., 1974) that expression of the genes encoding *E. coli* β-oxidation enzymes requires both cAMP and its receptor protein (CRP). Sequence analysis of the promoter region revealed that a 29-bp DNA fragment (from −48 to −76) has 8 of 12 nucleotides identical with the highly conserved base pairs of cAMP–CRP binding sites (de Crombrugghe et al., 1984). The distance between the center of the putative cAMP–CRP binding site and the transcription start site is −61/62 (Figure 2). Another CGTGA segment (from −81 to −85) may participate in the search of CRP for the primary binding site, but a secondary site is not essential for transcription activation (de Crombrugghe et al., 1984).

Analysis of the Multifunctional Protein Sequence. Since the activities of different mitochondrial β-oxidation enzymes are carried on separate polypeptide chains (Schulz, 1985), a comparison of the sequence of the large α-subunit of the *E. coli* fatty acid oxidation complex with the sequences of the corresponding mitochondrial β-oxidation enzymes was carried out in an effort to locate the various enzymatic functions within the multifunctional protein. As shown in Figure 4, the amino-terminal region (residues 4–253) is clearly homologous to rat mitochondrial enoyl-CoA hydratase (Minami-Ishii et al., 1989). The overall identity of the two peptides is 29%, and a region near the center (residues 108–148 in *E. coli* multifunctional protein and residues 104–144 of rat mitochondrial enoyl-CoA hydratase) shows 54% identity. This region is quite well conserved throughout the "hydratase family" (Minami-Ishii et al., 1989) and probably plays an important role in the catalytic function. On the basis of this evidence, we propose that the amino-terminal region of the *fadB* gene product is an enoyl-CoA hydratase functional domain. The enoyl-CoA hydratase has a broad chain-length specificity (Binstock & Schulz, 1981). The hydration of short-chain substrate (crotonyl-CoA) catalyzed by this enzyme is 2.5-fold faster than that of the medium-chain substrate (2-*trans*-decenoyl-CoA) (Binstock & Schulz, 1981). Clearly, this hydratase is not specific for medium-chain substrates, as suggested by DiRusso (1990).

A search for homology to pig heart mitochondrial L-3-hydroxyacyl-CoA dehydrogenase (Bitar et al., 1980) revealed clear homology with a central region (residues 294–593) of the *E. coli* multifunctional protein. In the alignment shown in Figure 4, there is a 36% identity between these two proteins. NAD^+ is a necessary cofactor for the dehydrogenase activity (Yang & Schulz, 1983), and by searching the sequence for the "fingerprint" consensus sequence of an ADP-binding βαβ-fold structure, a predicted NAD-binding domain was identified in this region. Since a 29 amino acid residue fragment (residues 315–343) has 10 of 11 residues that match the "fingerprint" and contains a perfect "core fingerprint" with glycine residues at positions, 6, 8, and 11 and with an acidic residue (aspartic acid) at the end (Wierenga et al., 1986), it is clearly homologous with the NAD-binding βαβ-fold of the pig heart mitochondrial L-3-hydroxyacyl-CoA dehydrogenase

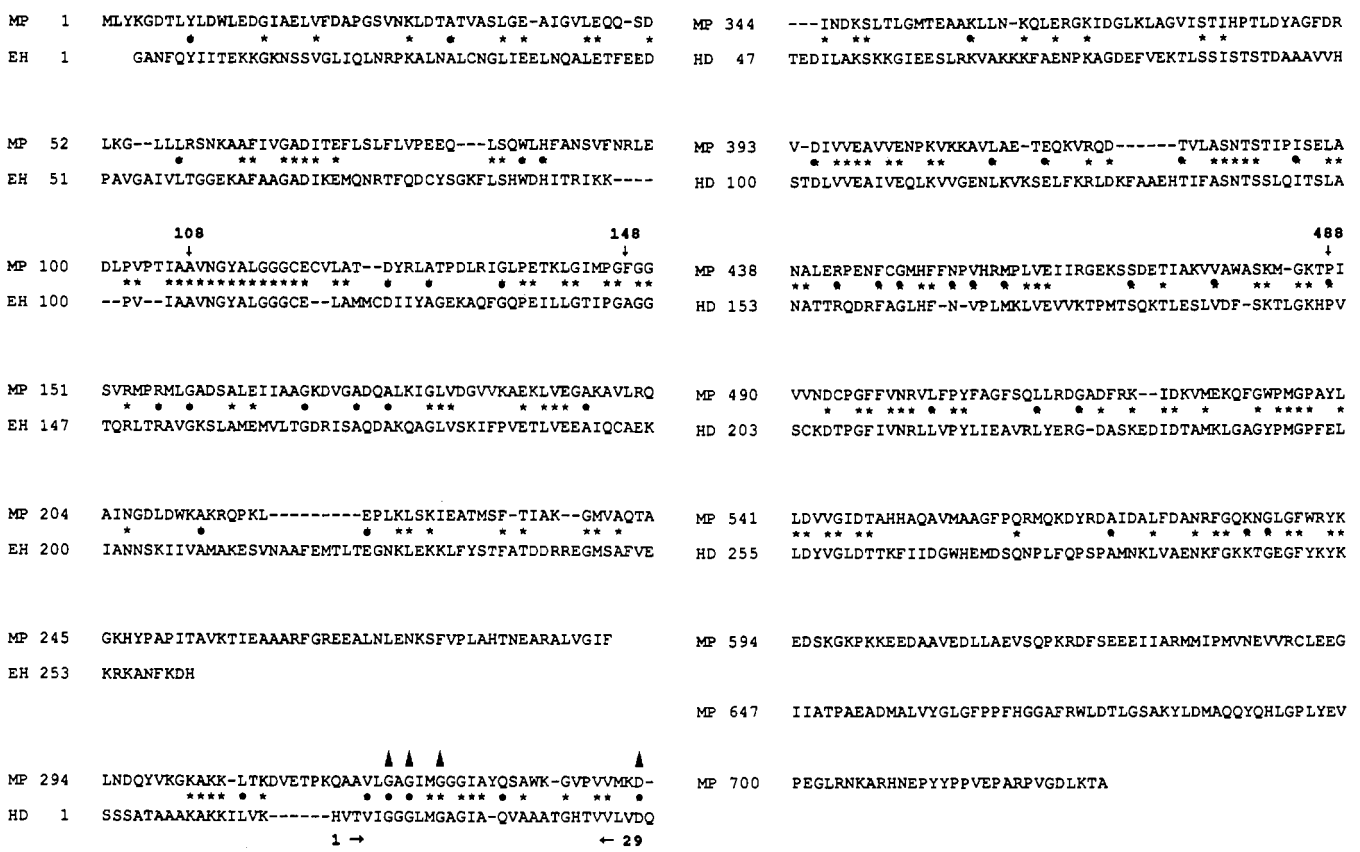

FIGURE 4: Comparison of the amino acid sequence of the *E. coli* multifunctional protein (MP) with those of rat mitochondrial enoyl-CoA hydratase (EH) and pig heart mitochondrial L-3-hydroxyacyl-CoA dehydrogenase (HD). Standard one-letter amino acid abbreviations are used. Pairs of sequences are aligned with gaps (-) inserted to obtain maximal similarities. Residues identical in two sequences are shown by asterisks (★). Solid triangles indicate the positions of the "core fingerprint", and the numbers below the sequence indicate the positions important to the formation of a $\beta\alpha\beta$-fold, which has been shown to bind the ADP moiety of NAD (Wierenga et al., 1986). The sequences of rat mitochondrial enoyl-CoA hydratase and pig heart mitochondrial L-3-hydroxyacyl-CoA dehydrogenase are taken from Minami-Ishii et al. (1989) and Bitar et al. (1980), respectively.

(see Figure 4). The conformation of the latter enzyme has been determined by X-ray diffraction (Birktoft et al., 1987); it has a two-domain structure with a bilobal appearance. The larger amino-terminal lobe (residues 1–200) binds the NAD cofactor and displays considerable structural homology with the NAD-binding domains observed in other dehydrogenases. The L-3-hydroxyacyl-CoA substrate binds in the cleft between the lobes. The carboxyl-terminal smaller lobe bears little homology to other known dehydrogenases and forms the subunit–subunit interface in the dimer (Birktoft et al., 1987). As shown in Figure 4, the sequence homology between mitochondrial and *E. coli* L-3-hydroxyacyl-CoA dehydrogenase extends over a range of about 300 residues. A fragment of the *E. coli* multifunctional protein (residues 294–487) corresponding to the amino-terminal domain (residues 1–200) of the mitochondrial dehydrogenase contains the putative NAD-binding $\beta\alpha\beta$-fold and must be responsible for binding the NAD cofactor. The next 106-residue fragment (residues 488–593) of the *fadB* gene product is very similar to the carboxyl-terminal smaller domain (residues 201–307) of the pig heart dehydrogenase. This finding also suggests that the gene of the latter enzyme probably evolved by partial duplication of an ancient multifunctional protein gene.

Δ^3-*cis*-Δ^2-*trans*-Enoyl-CoA isomerase has recently been purified from rat liver mitochondria (Palosaari et al., 1990). However, the amino acid sequence of this enzyme has not yet been determined. It seems likely to us that Δ^3-*cis*-Δ^2-*trans*-enoyl-CoA isomerase is present in the carboxyl-terminal region (residues 594–729) of the multifunctional protein because the rest of the protein already harbors other enzymes (Figure 4). Since the chemical modification of the multifunctional protein by *N*-ethylmaleimide only inhibited the isomerase and did not affect other enzyme activities (Pawar & Schulz, 1981), the functional domain of the isomerase is probably physically distinct from the others. The exact location and size of the isomerase functional domain, of course, need to be further confirmed by protein engineering, X-ray crystallography, or other techniques.

Besides the hydratase, dehydrogenase, and isomerase activities mentioned above, this multifunctional polypeptide also possesses the ability to epimerize D-3-hydroxyacyl-CoAs to their L-isomers (Yang & Shulz, 1983; Yang et al., 1986a). Recent studies have revealed that the epimerization of D-3-hydroxyacyl-CoAs in rat liver peroxisomes is accomplished by a two-step dehydration–hydration reaction sequence that is catalyzed by a novel hydratase and crotonase (Smeland et al., 1989; Hiltunen et al., 1989; Li et al., 1990). The epimerization of D-3-hydroxyacyl-CoAs in *E. coli* probably proceeds by the same mechanism, not through redox catalysis using NAD as a cofactor as seen in UDP-glucose 4-epimerase (Glaser, 1972; Mukherji & Bhaduri, 1986), because the following observations suggest that the *E. coli* L-3-hydroxyacyl-CoA dehydrogenase does not function as the epimerase: (a) while the 3-hydroxyacyl-CoA dehydrogenase activity was strongly inhibited by iodoacetamide labeling of the NAD binding site, the epimerase activity was only slightly affected by this reaction (Yang & Schulz, 1983), and (b) a mutant of the *E. coli* fatty acid oxidation complex that exhibited a normal level of epimerase activity, but lacked L-3-hydroxyacyl-CoA dehydrogenase activity, has been purified (Pramanik & Schulz,

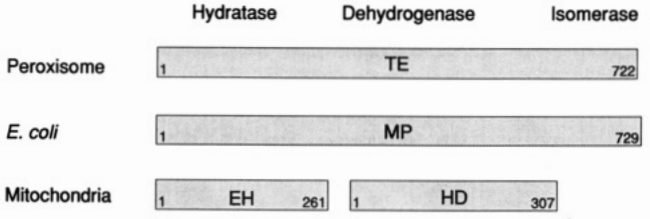

FIGURE 5: Locations of functional domains in *E. coli* multifunctional protein (MP) and rat peroxisomal trifunctional β-oxidation enzyme (TE). The amino-terminal and the central regions of these two proteins are homologous to mitochondrial enoyl-CoA hydratase (EH) and L-3-hydroxyacyl-CoA dehydrogenase (HD), respectively. The rat peroxisomal trifunctional β-oxidation enzyme does not exhibit 3-hydroacyl-CoA epimerase activity, and we propose that the conserved carboxyl-teminal domains of these two proteins harbor Δ^3-*cis*-Δ^2-*trans*-enoyl-CoA isomerase. The epimerase activity of the *E. coli* multifunctional protein is presumably associated with the enoyl-CoA hydratase functional domain.

1983). The 3-hydroxyacyl-CoA epimerase activity is thus probably associated with the amino-terminal hydratase domain of the *E. coli* multifunctional protein. In summary, on the basis of sequence homologies we propose that the enzymatic activities associated with the multifunctional protein are located in the order: enoyl-CoA hydratases (3-hydroxyacyl-CoA epimerase and crotonase):L-3-hydroxyacyl-CoA dehydrogenase:Δ^3-*cis*-Δ^2-*trans*-enoyl-CoA isomerase (see Figure 5).

Sequence Similarity between Rat Peroxisomal Trifunctional β-Oxidation Enzyme and the E. coli Multifunctional Protein. Rat enoyl-CoA hydratase:L-3-hydroxyacyl-CoA dehydrogenase:Δ^3-*cis*-Δ^2-*trans*-enoyl-CoA isomerase trifunctional enzyme catalyzes the second and third reactions of the peroxisomal β-oxidation system and additionally the isomerization of 3-enoyl-CoA thioesters (Palosaari & Hiltunen, 1990). It has been reported (Ishii et al., 1987; Minami-Ishii et al., 1989) that the amino-terminal region (residues 1–188 encoded by exons I–V) and the fragment from residues 296 to 527 of the trifunctional enzyme are homologous to mitochondrial enoyl-CoA hydratase and 3-hydroxyacyl-CoA dehydrogenase, respectively. The similarity terminates abruptly at the end of exon V, and there is essentially no similarity between the polypeptide fragment encoded by exon VI (residues 189–302) and the corresponding regions of mitochondrial enzymes (Ishii et al., 1987; Minami-Ishii et al., 1989). In contrast, the sequence similarity between the *E. coli* multifunctional protein and rat peroxisomal trifunctional enzyme is apparent over the full lengths of the polypeptides including the region encoded by exon VI of the gene for rat peroxisomal trifunctional enzyme (Figure 6). As shown in Figure 6, the carboxyl-terminal regions of these two proteins are definitely related, as judged by the criteria published by Dayhoff et al. (1983). The similarities are greatest in a segment of the polypeptide (residues 611–661 of rat peroxisomal trifunctional enzyme) where 21 out of 51 residues are identical with those in a counterpart of the *E. coli* multifunctional protein without the introduction of breaks or gaps in either sequence. The carboxyl-terminal region of rat peroxisomal trifunctional β-oxidation enzyme (residues 581–722) presumably has the same function as the corresponding region of the *E. coli* multifunctional protein, which we propose to be the Δ^3-*cis*-Δ^2-*trans*-enoyl-CoA isomerase activity (see Figure 5).

The rat peroxisomal trifunctional β-oxidation enzyme differs from the *E. coli* multifunctional protein in that it does not exhibit 3-hydroxyacyl-CoA epimerase activity (Yang &

```
TE   1   MAEYLRLPHSLAMIRLCNPP--VNAV-SPTVIREV-RNGLQKAGSDHT
MP   1   MLYKGDTLYLDWLEDGIAELDFDAPGSVNKLDTATVASLGEAIGVLEQQSD--

TE  45   VKAIVICGANGNFCAGADIHGF-SAFTPGLALGSLVDEIQR------YQKPVL
MP  52   LKGLLLRSNKAAFIVGADITEFLSLFLVPEEQLSQWLHFANSVFNRLEDLPVP

TE  91   --AAIQGVALGGGLELALGCHYRIANAKARVGLPEVTLGILPGARGTQLLPRV
MP 105   TIAAVNGYALGGGCECVLATDYRLATPDLRIGLPETKLGIMPGFGGSVRMPRM

TE 142   VGVPVALDLITSGKYLSADEALRLGILDAVVKSD-PVEEAIKFA--QKI-IDK
MP 158   LGADSALEIIAAGKDVGADQALKIGLVDGVVKAEKLVEGA-KAVLRQAINGDL

TE 191   PIEPRRIFNKPVPSLPNMDSVFAEAIAK-VRKQYPGVL--APETCVRSIQASV
MP 210   DWKAKRQPKLEPLKLSKIEATMSFTIAKGMVAQTAGKHYPAPITAVKTIEAAA

TE 241   KHPYEVGIKEEEKLFMYLRASGQAKALQYAFFAEKSAQKWSTPSGASWK-TAS
MP 263   RFGREEALNLENKSFVPLAHTNEARALVGIF------LNDQYVKGKAKKLTKD

TE 293   AQPVSSVGVLGLGTMGRGIAI-SFARVGISVVAVESDPKQLDAAKKIITFTLE
MP 310   VETPKQAAVLGAGIMGGGIAYQS-AWKGVPVVMKDINDKSLT-----LGMTEA

TE 345   KEASRAHQNGQASAKPKLRFSSSTKELST-------VDLVVEAVFEDMNLK-K
MP 357   AKLLNKQLERGKIDGLKLAGVISTIHPTLDYAGFDRVDIVVEAVVE--NPKVK

TE 390   K-VFAELSALCKPGAFLCTNTSALNVDDIASSTDRPQLVIGTHFFSPAHVMRL
MP 408   KAVLAETEQKVRQDTVLASNTSTIPISELANALERPENFCGMHFFNPVHRMPL

TE 442   LEVIPSRYSSPTTIATVMSL-SKKIGKIGVVVGN-CYGF-VGNRMLAPYYNQG
MP 461   VEIIRGEKSSDETIAKVVAWASKM-GKTPIVV-NDCPGFFV-NRVLFPYFA-G

TE 492   FF-LLEEGSKPEDVDGVLEE-FGFKMGPFRVSDLAGLDVGWKIRKGQGLTGPS
MP 510   FSQLLRDGADFRKIDKVMEKQFGWPMGPAYLLDVVGIDTAHHAQAVMAA----

                                                    580
                                                     ↓
TE 543   LPPGTPVRKRGNSRYSPLGDMLCEAGRFGQKTGKG-W-YQYD-KPLGRIHKP-D
MP 559   ---GFPQRMQKD--YRDAIDALFDANRFGQKNGLGFWRYKEDSK--GKPKKEED

                                          611
                                           ↓
TE 593   PWLSTFLSQYREVHHIEQRTISKEEILERCLYSLINEAFRILEEGMAARPEHID
MP 606   AAVEDLL---AEVSQP-KRDFSEEEIIARMMIPMVNEVVRCLEEGIIATPAEAD

                                    661
                                     ↓
TE 647   VIYLHGYGWPRHKGGPMFYAASV-GLPTVLEKLQKYYRQNPDIPQLEPSDYLRR
MP 656   MALVYGLGFPPFHGGA-FRWLDTLGSAKYLDMAQQYQHLGPLYEVPEGLRNKAR

TE 700   LVAQGSPPLKEWQSLAGPHGSKL
MP 709   HNEPYYPPV-EPARPVGDLKTA
```

FIGURE 6: Comparison between the amino acid sequences of rat peroxisomal trifunctional β-oxidation enzyme (TE) and *E. coli* multifunctional protein (MP). Standard one-letter amino acid abbreviations are used. The two sequences are aligned with appropriate gaps (–) to obtain maximum homology. Residues identical in both sequences are shown by asterisks (★). The rat peroxisomal trifunctional β-oxidation enzyme gene is divided into seven exons by six introns. Downward triangles indicate the positions of intron insertion. The sequence of rat peroxisomal trifunctional β-oxidation enzyme is taken from Osumi et al. (1985).

Schulz, 1983; Palosaari et al., 1990). However, the sequence of the dehydrogenase functional domain of the rat peroxisomal trifunctional enzyme (residues 278–580) shows greater homology to the corresponding dehydrogenase functional domain of the *E. coli* multifunctional protein (residues 294–593) (Figure 6) than to the pig heart mitochondrial L-3-hydroxyacyl-CoA dehydrogenase (Ishii et al., 1987). Moreover, the sizes of the rat peroxisomal trifunctional enzyme (78 kDa) and of the *E. coli* multifunctional protein (79 kDa) are almost identical, and the hydratase and dehydrogenase in both proteins (Figure 5) are capable of catalyzing the direct β-oxidation of 2,4-dienoyl-CoAs formed during the β-oxidation of unsaturated fatty acids (Yang et al., 1986a,b). This metabolic feature may be due to juxtaposition of the functional domains having these enzymatic activities in nonmitochondrial β-oxidation systems. All together, the evidence suggests that the genes of the rat peroxisomal trifunctional β-oxidation enzyme and the *E. coli* multifunctional protein have a common ancestor. This is consistent with the proposal (Yang et al., 1990) that the peroxisomal 3-ketoacyl-CoA thiolase gene evolved directly from the prokaryotic 3-ketoacyl-CoA thiolase gene. The mammalian peroxisomal β-oxidation system thus appears to be evolutionarily more closely related to the prokaryotic β-oxidation system than to the mitochondrial β-oxidation system, and this provides evidence for the hypothesis (de Duve, 1983) that peroxisomes are descendants of ancient endosymbionts.

References

Bachmann, B. J. (1990) *Microbiol. Rev. 54*, 130–197.

Binstock, J. F., & Schulz, H. (1981) *Methods Enzymol. 71*, 403–411.

Birktoft, J. J., Holden, H. M., Hamlin, R., Xuong, N. H., & Banaszak, L. J. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 8262–8266.

Bitar, K. G., Perez-Aranda, A., & Bradshaw, R. A. (1980) *FEBS Lett. 116*, 196–198.

Caruthers, M. H. (1985) *Science 230*, 281–285.

Chaconas, G., & van de Sande, J. H. (1980) *Methods Enzymol. 65*, 75–88.

Chung, C. T., Niemela, S. L., & Miller, R. H. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 2172–2175.

Dale, R. M. K., McClure, B. A., & Houchins, J. P. (1985) *Plasmid 13*, 31–40.

Dayhoff, M. O., Barker, W. C., & Hunt, L. T. (1983) *Methods Enzymol. 91*, 524–545.

Debarbouille, M., & Raibaud, O. (1983) *J. Bacteriol. 153*, 1221–1227.

de Crombrugghe, B., Busby, S., & Buc, H. (1984) *Science 224*, 831–838.

de Duve, C. (1983) *Sci. Am. 248*, 74–84.

DiRusso, C. C. (1990) *J. Bacteriol. 172*, 6459–6468.

Gilman, M. Z., & Chamberlin, M. J. (1983) *Cell 35*, 285–293.

Glaser, L. (1972) in *The Enzymes* (Boyer, P. D., Ed.) 3rd ed., Vol. 6, pp 355–380, Academic Press, Orlando, FL.

Hanahan, D. (1983) *J. Mol. Biol. 166*, 557–580.

Hiltunen, J. K., Palosaari, P. M., & Kunau, W.-H. (1989) *J. Biol. Chem. 264*, 13536–13540.

Howarth, A. J., Gardner, R. C. Messing, J., & Sheperd, R. J. (1981) *Virology 112*, 678–685.

Innis, M. A., Myambo, K. B., Gelfand, D. H., & Brow, M. A. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 9436–9440.

Ishii, N., Hijikata, M., Osumi, T., & Hashimoto, T. (1987) *J. Biol. Chem. 262*, 8144–8150.

Li, J., Smeland, T. E., & Schulz, H. (1990) *J. Biol. Chem. 265*, 13629–13634.

Minami-Ishii, N., Taketani, S., Osumi, T., & Hashimoto, T. (1989) *Eur. J. Biochem. 185*, 73–78.

Mizusawa, S., Nishimura, S., & Seela, F. (1986) *Nucleic Acids Res. 14*, 1319–1324.

Moore, J., Engelberg, A., & Bairoch, A. (1988) *BioTechniques 6*, 566–572.

Mukherji, S., & Bhaduri, A. (1986) *J. Biol. Chem. 261*, 4519–4524.

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

Narnes, W. M., Bevan, M., & Son, P. H. (1983) *Methods Enzymol. 101*, 98–122.

Nunn, W. D. (1986) *Microbiol. Rev. 50*, 179–192.

Osumi, T., Ishii, N., Hijikata, M., Kamijo, K., Ozasa, H., Furuta, S., Miyazawa, S., Kondo, K., Inoue, K., Kagamiyama, H., & Hashimoto, T. (1985) *J. Biol. Chem. 260*, 8905–8910.

Overath, P., Pauli, G., & Schairer, H. U. (1969) *Eur. J. Biochem. 7*, 559–574.

Palosaari, P. M., & Hiltunen, J. K. (1990) *J. Biol. Chem. 265*, 2446–2449.

Palosaari, P. M., Kilponen, J. M., Sormunen, R. T., Hassinen, I. E., & Hiltunen, J. K. (1990) *J. Biol. Chem. 265*, 3347–3353.

Pauli, G., Ehring, R., & Overath, P. (1974) *J. Bacteriol. 117*, 1178–1183.

Pawar, S., & Schulz, H. (1981) *J. Biol. Chem. 256*, 3894–3899.

Pramanik, A., & Schulz, H. (1983) *Biochim. Biophys. Acta 750*, 41–46.

Raibaud, O., & Schwartz, M. (1984) *Annu. Rev. Genet. 18*, 173–206.

Rosenberg, M., & Court, D. (1979) *Annu. Rev. Genet. 13*, 19–53.

Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463–5467.

Schulz, H. (1985) in *Biochemistry of Lipids and Membranes*, (Vance, D. E., & Vance, J. E., Eds.) pp 116–142, Benjamin/Cummings, Menlo Park, CA.

Smeland, T. E., Li, J., Chu, C.-H., Cuebas, D., & Schulz, H. (1989) *Biochem. Biophys. Res. Commun. 160*, 988–992.

Smith, L. M., Sanders, J. Z., Kaiser, R. J., Hughes, P., Dodd, C., Connell, C. R., Heiner, C., Kent, S. B. H., & Hood, L. E. (1986) *Nature 321*, 674–679.

Stoffel, W., & Caesar, H. (1965) *Hoppe-Seyler's Z. Physiol. Chem. 341*, 76–83.

Summers, W. C. (1970) *Anal. Biochem. 33*, 459–463.

Volgelstein, B., & Gillespie, D. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 615–619.

Wierenga, R. K., Terpstra, P., & Hol, W. G. J. (1986) *J. Mol. Biol. 187*, 101–107.

Williams, J. G., & Mason, P. J. (1985) in *Nucleic Acid Hybridization* (Hames, B. D., & Higgins, S. J., Eds.) pp 139–160, IRL Press, Oxford.

Yang, S.-Y., & Schulz, H. (1983) *J. Biol. Chem. 258*, 9780–9785.

Yang, S.-Y., Bittman, R., & Schulz, H. (1985) *J. Biol. Chem. 260*, 2862–2868.

Yang, S.-Y., Cuebas, D., & Schulz, H. (1986a) *J. Biol. Chem. 261*, 12238–12243.

Yang, S.-Y., Cuebas, D., & Schulz, H. (1986b) J. Biol. Chem. 261, 15390–15395.
Yang, S.-Y., Li, J., He, X.-Y., Cosloy, S. D., & Schulz, H. (1988) J. Bacteriol. 170, 2543–2548.
Yang, S.-Y., Yang, X.-Y. H., Healy-Louie, G., Schulz, H., & Elzinga, M. (1990) J. Biol. Chem. 265, 10424–10429.
Yanisch-Perron, C., Vierra, J., & Messing, J. (1985) Gene 33, 103–119.

Affinity Labeling of the Active Site and the Reactive Sulfhydryl Associated with Activation of Rat Liver Phenylalanine Hydroxylase†

Barbara Schuster Gibbs and Stephen J. Benkovic*
Department of Chemistry, The Pennsylvania State University, University Park, Pennsylvania 16802
Received November 7, 1990; Revised Manuscript Received April 9, 1991

ABSTRACT: A pterin analogue, 5-[(3-azido-6-nitrobenzylidene)amino]-2,6-diamino-4-pyrimidinone (ANBADP), was synthesized as a probe of the pterin binding site of phenylalanine hydroxylase. The photoaffinity label has been found to be a competitive inhibitor of the enzyme with respect to 6,7-dimethyltetrahydropterin, having a K_i of 8.8 ± 1.1 μM. The irreversible labeling of phenylalanine hydroxylase by the photoaffinity label upon irradiation is both concentration and time dependent. Phenylalanine hydroxylase is covalently labeled with a stoichiometry of 0.87 ± 0.08 mol of label/enzyme subunit. 5-Deaza-6-methyltetrahydropterin protects against inactivation and both 5-deaza-6-methyltetrahydropterin and 6-methyltetrahydropterin protect against covalent labeling, indicating that labeling occurs at the pterin binding site. Three tryptic peptides were isolated from [^{3}H]ANBADP-photolabeled enzyme and sequenced. All peptides indicated the sequence Thr-Leu-Lys-Ala-Leu-Tyr-Lys (residues 192–198). The residues labeled with [^{3}H]ANBADP were Lys198 and Lys194, with the majority of the radioactivity being associated with Lys198. The reactive sulfhydryl of phenylalanine hydroxylase associated with activation of the enzyme was also identified by labeling with the chromophoric label 5-(iodoacetamido)fluorescein [Parniak, M. A., & Kaufman, S. (1981) J. Biol. Chem. 256, 6876]. Labeling of the enzyme resulted in 1 mol of fluorescein bound per phenylalanine hydroxylase subunit and a concomitant activation of phenylalanine hydroxylase to 82% of the activity found with phenylalanine-activated enzyme. Tryptic and chymotryptic peptides were isolated from fluorescein-labeled enzyme and sequenced. The modified residue was identified as Cys236.

Phenylalanine hydroxylase (EC 1.14.16.1) (PAH)[1] catalyzes the hydroxylation of phenylalanine to tyrosine in the presence of molecular oxygen. PAH also shows an absolute requirement for a tetrahydropterin (Kaufman, 1959). Although tetrahydrobiopterin is the natural substrate (Kaufman, 1963) a number of synthetic pterins such as 6-MPH$_4$, 6,7-DMPH$_4$, and 6-phenyltetrahydropterin can be used as cofactors for PAH (Bailey & Ayling, 1978; Kaufman & Levenberg, 1959). A pyrimidine analogue of 6-phenyltetrahydropterin, 5-(benzylamino)-2,6-diamino-4-pyrimidinone is also active as a cofactor for PAH (Bailey & Ayling, 1978, 1980). It binds with the same affinity as 6-phenyltetrahydropterin but has a maximum velocity only 1.5% that of the analogous pterin.

Although much is known about the pterin cofactor requirements, nothing is known about the pterin binding site. There is little structural information available regarding PAH or any of its binding sites. For this reason an azido analogue of 6-phenyltetrahydropterin has been synthesized in this laboratory to probe the pterin binding site of PAH. The photoaffinity label 5-[(3-azido-6-nitrobenzylidene)amino]-2,6-diamino-4-pyrimidinone (ANBADP) is an analogue of 6-phenyldihydropterin lacking the 7-carbon of the pteridine ring and containing azido and nitro substituents on the phenyl ring. We report here the interaction of this photoaffinity label with PAH, its covalent binding to the enzyme upon photolysis, and the isolation and sequence of a labeled peptide.

Rat liver PAH is also under allosteric control and is reversibly activated by its substrate, phenylalanine (Shiman et al., 1979, 1990; Shiman & Gray, 1980). This allosteric binding site is a unique site and separate from the catalytic site (Shiman et al., 1990; Shiman, 1980). In rat liver PAH there is one free exposed sulfhydryl in the native enzyme (Fisher et al., 1972), and modification of this sulfhydryl with *N*-ethylmaleimide results in activation of the enzyme (Parniak & Kaufman, 1981). A concomitant apparent decrease in total phenylalanine binding suggests that this sulfhydryl is at or adjacent to the allosteric activation site. In order to understand more about this site, we have labeled the free sulfhydryl with the chromophore 5-(iodoacetamido)fluorescein (IAF). We report here the isolation and sequence of fluorescein-labeled peptides isolated from proteolytic digestion of labeled native PAH.

Experimental Procedures

Materials. Manganese dioxide, 3-fluorobenzaldehyde, sodium borodeuteride, and 2,5,6-triamino-4-pyrimidinone

† This work was supported by NSF Grant DMB-8614179.
* Author to whom correspondence should be addressed.

[1] Abbreviations: PAH, phenylalanine hydroxylase; ANBADP, 5-[(3-azido-6-nitrobenzylidene)amino]-2,6-diamino-4-pyrimidinone; FNBADP, 5-[(3-fluoro-6-nitrobenzylidene)amino]-2,6-diamino-4-pyrimidinone; 5-IAF, 5-(iodoacetamido)fluorescein; 6-MPH$_4$, 6-methyl-5,6,7,8-tetrahydropterin; 6,7-DMPH$_4$, 6,7-dimethyl-5,6,7,8-tetrahydropterin; DTT, dithiothreitol; HPLC, high-pressure liquid chromatography; TFA, trifluoroacetic acid.

0006-2960/91/0430-6795$02.50/0 © 1991 American Chemical Society