Volume 16 Number 16 1988 Nucleic Acids Research

Nucleotide sequence of the *fadR* gene, a multifunctional regulator of fatty acid metabolism in *Escherichia coli*

Concetta C.DiRusso

College of Medicine, Department of Biochemistry, University of Tennessee, 800 Madison Avenue, Memphis, TN 38163, USA

Received April 7, 1988; Revised and Accepted July 18, 1988 Accession no. X08087

ABSTRACT

The *Escherichia coli fadR* gene is a multifunctional regulator of fatty acid and acetate metabolism. In the present work the nucleotide sequence of the 1.3 kb DNA fragment which encodes FadR has been determined. The coding sequence of the *fadR* gene is 714 nucleotides long and is preceded by a typical *E. coli* ribosome binding site and is followed by a sequence predicted to be sufficient for factor-independent chain termination. Primer extension experiments demonstrated that the transcription of the *fadR* gene initiates with an adenine nucleotide 33 nucleotides upstream from the predicted start of translation. The derived *fadR* peptide has a calculated molecular weight of 26,972. This is in reasonable agreement with the apparent molecular weight of 29,000 previously estimated on the basis of maxi-cell analysis of plasmid encoded proteins. There is a segment of twenty amino acids within the predicted peptide which resembles the DNA recognition and binding site of many transcriptional regulatory proteins.

INTRODUCTION

Upon growth of *E. coli* cells in minimal medium containing long-chain fatty acids (C_{12}-$C_{18:1}$), the products of the *fadL* (50 min), *fadD* (24 min), *fadE* (5 min), *fadAB* (85 min) and *aceBAK* (90 min) genes and operons are coordinately induced (reviewed in reference 1). These hydrophobic compounds enter the cell by a protein-mediated translocation step involving the products of at least the *fadL* and *fadD* genes (2,3,4). The product of the *fadL* gene is located on the outer membrane where it appears to act as a receptor for long-chain fatty acids (5). The *fadD* gene encodes fatty acyl-CoA synthetase (6). Due to the coupled transport and activation processes, the main pool of long-chain fatty acids within the cell is in the form of fatty acyl-CoA rather than free fatty acids(7). The fatty acyl-CoA's (C_6-$C_{18:1}$) may be degraded via ß-oxidation and thiolytic cleavage or may be converted to phosphatidic acid to be incorporated directly into phospholipids (8). At least six additional enzyme activities (the

products of the *fadE*, *fadA*, and *fadB* genes) are required to carry out these reactions which result in the release of one molecule of acetyl-CoA for each turn of the cycle (6-11).

Acetyl-CoA produced by ß-oxidation of fatty acids can serve as a substrate for *de novo* synthesis of fatty acids and phospholipids or can be metabolized via the tricarboxylic acid cycle to serve as a carbon and energy source (12). Growth on fatty acids or acetate as a sole carbon and energy source also requires the enzymes of the glyoxylate shunt which are encoded within the *aceBAK* operon (13,14). These enzymes allow the net assimilation of carbon from acetyl-CoA to replenish the dicarboxylic acids necessary for the biosynthesis of cellular components. The biosynthesis of fatty acids occurs by the condensation of acetyl-CoA with malonyl- CoA and subsequent chain elongation by the repeated addition of malonyl-CoA by the fatty acid synthetase complex and acyl carrier protein to form a ß-keto fatty acid derivative (8). The fatty acids derived by *de novo* synthesis may be incorporated into phospholipids, glycolipids and lipoproteins (8). Thus the enzymatic pathways of fatty acid degradation, acetate degradation and fatty acid biosynthesis are linked biochemically through several common intermediates including acetyl-CoA.

In addition to the biochemical links detailed above, fatty acid degradation fatty acid biosynthesis and acetate metabolism are coordinately regulated in *E. coli* by the product of the *fadR* gene, FadR (6). In fact, the FadR system appears to form a global control network as described by Gottesman for the cAMP-CRP system, the SOS system and the proteins involved in the heat shock response (15). Spontaneous mutations in *fadR* cause the constitutive expression of the *fad* structural genes. These mutants are isolated at a fequency of $1/10^6$ by growth on minimal medium containing medium chain fatty acids as the sole carbon and energy source and map to 25.5 minutes on the *E. coli* genetic map (6,16). Medium-chain fatty acids are degraded by the same enzymatic machinery as long-chain fatty acids but do not serve as a carbon source. Transposon insertions in *fadR* which cause highly polar effects and abolish gene expression also result in constitutive levels of the *fadL*, *fadD*, *fadE*, *fadAB* and *aceBAK* enzymes (1, 17, 18). $fadR^+$ supplied in *trans* on a F' episome (26) or a plasmid (1) is dominant to the mutant *fadR* in restoring the inducible phenotype to the mutant strain. This information suggests that *fadR* is a negative regulator of the *fad* and *ace* regulon and that a long-chain fatty acid or a long-chain fatty acid derivative is the inducer.

Further evidence suggesting that *fadR* controls the *fad* and *ace* genes at the level of transcription was provided by studies using operon fusions which place the structural gene for ß-galactosidase, *lacZ*, under the control of the *fad* or *ace* regulatory regions and promoters. *fadR* strains harboring Φ(*fadL-lacZ*) (20), Φ(*fadE-lacZ*) (19, 21), Φ(*fadAB-lacZ*) (21), Φ(*aceB-lacZ*) (22), or Φ(*aceA-lacZ*) (22) have elevated levels of ß-galactosidase activity as compared with *fadR*$^+$ strains whether or not long chain-fatty acids (and/or acetate in the case of *ace-lacZ* fusions) are present in the growth medium. *fadR*$^+$ strains harboring the fusions have inducible levels of ß-galactosidase. Therefore, FadR appears to be a *trans*- acting protein which controls fatty acid degradation and acetate degradation by decreasing the expression of the *fad* and *ace* structural genes. The regulation of the *ace* genes also involves the gene *ic1R*. Strains carrying the *ace-lacZ* fusions and a mutation in *ic1R* have elevated levels of ß-galactosidase (22). Strains which carry both mutations, *fadR* and *ic1R*, have ß-galactosidase levels higher than strains with either mutation alone implying these regulators have an independent and additive effect (18, 22). Therefore, transcriptional control of acetate metabolism by FadR and Ic1R is predicted to be complex.

The role of FadR in fatty acid biosynthesis is less well defined than its established role in the control of transcription of the *fad* and *ace* genes. Strains which carry a mutation in *fadR* are elongated as compared to wild-type strains when grown under conditions which favor rapid growth (*i.e.* LB at 37^{o}C) (23). The *fadR* cells often fail to septate and long filaments are formed. This morphological abnormality, which resembles that of the *fts* and *env* mutants (24), is accompanied by a change in the phospholipid composition. *fadR* mutants synthesize significantly less unsaturated fatty acids than do wild-type cells (25). This alteration is phenotypically asymptomatic unless the *fadR* strain also carries a lesion in *fabA*, the structural gene for ß-hydroxydecanoyl-thioester dehydrase (26, 27). This enzyme adds the *cis*-double bond at the $C_{10:0}$ stage to the elongating fatty acid chain and is required for unsaturated fatty acid biosynthesis. Unlike *fadR*$^+$ *fabA*(Ts) mutants, *fadR fabA*(Ts) strains synthesize insufficient levels of unsaturated fatty acids to sustain growth at both low and high temperatures and require unsaturated fatty acid supplementation at all temperatures (27). The affect of a mutation in *fadR* on fatty acid levels is not a result of increased ß-oxidation due to constitutive levels of the *fad* enzymes since mutations in the *fad* structural genes do not alleviate these affects (*fadD*,

fadE and *fadAB* were tested (27)). These results suggest that *fadR* positively regulates unsaturated fatty acid biosynthesis at the level of *fabA* expression. However, the pleiotropic affects of *fadR* on phospholipid and unsaturated fatty acid biosynthesis may be an indirect result of modulation by a secondary effector(s) itself controlled by $fadR^+$. We cannot distinguish between these possibilities at this time.

To further define the role of *fadR* in the regulation of these fatty acid and lipid-related pathways, we have previously cloned and preliminarily characterized *fadR* (19). A single polypeptide (29,000M_r) specifically associated with wild-type *fadR* gene activity was identified by maxi-cell analysis of plasmid encoded proteins (19). In this work, I present the complete nucleotide sequence of the *fadR* gene and mapping of the transcription initiation site. A single open reading frame capable of encoding *fadR* was derived from the DNA sequence. This derived protein sequence is believed on the basis of the transcription analysis and deletion analysis to be the *fadR* peptide sequence.

METHODS

Bacterial strains and Media.

The *E. coli* host JM103 was used for M13 propagation (28). Methods of phage growth and DNA isolation were as described by J. Messing (28).

For analysis of the *fadR* phenotype strain LS1085 (*fadR*::Tn*10* (19)) was transformed with the plasmid of interest using standard $CaCl_2$ treatment procedures(29). Transformants were selected on solid LB medium (30) containing 100μg/ml ampicillin. Transformants were stippled to solid medium E (31) containing 0.01% leucine and 0.01% threonine and one of the following carbon sources: 50mM acetate; 5mM oleate in 0.5% Brij 58; or 5mM decanoate in 0.5% Brij 58. Growth on minimal medium containing decanoate ($C_{10:0}$) is indicative of the *fadR* mutation (6). The *fadR* phenotype conferred by these plasmids was also evaluated by transformation of LS1155, *fadR*::Tn*10*/ λΦ*fadA-lacZ*. This strain was generated as follows. The strain DC530 which carries *fadA*::Mu *d* (Ap, *lacZ*)(21) was stabilized with λpL209 as described by Komeda and Ino (32) to generate strain LS1154. LS1154 was made *fadR*::Tn*10* by generalized transduction with P1*vir* grown on LS5381(16) with primary selection for tetracycline resistance encoded within Tn*10* to give LS1155. LS1155 was transformed with the plasmid of interest as described in the text

with selection of transformants on MacConkey indicator plates containing 1% lactose and 100μg/ml ampicillin. Strain LS1154 (*fadR*$^+$)is white on these plates while LS1155 (*fadR*) is red due to repression of ß-galactosidase activity of Φ*fadA-lacZ* by the trans-acting product of the *fadR* gene.

Construction of plasmids containing deletions of *fadR* sequences.

To generate CD1901, the 1.3kb *Hin*DIII-*Eco*RV fragment of *E. coli* chromosomal DNA carrying the entire *fadR* gene (19) was cloned into *Hin*DIII-*Hin*cII restricted M13mp19 (28). Clones CD1902-1907 as illustrated in Figure 1 carrying progressive deletions were generated by restricting CD1901 with the endonucleases *Sst*I and *Xba*I followed by *Exo*III digestion as described by S. Henikoff (33). The clone CD2001 containing the *fadR* gene in the opposite orientation to CD1901 was generated by filling in the *Hin*DIII end of the same DNA fragment encoding *fadR* with deoxynucleotides using the Klenow fragment of DNA polymerase I, ligating this fragment to *Hin*cII restricted M13mp19, and identification of the clone of interest by the C-test (28). Deletions of CD2001 (CD2002-CD2007 as illustrated in Figure 1) were generated in the same manner as those of CD1901.

Determination of nucleotide sequence.

DNA sequence analysis was performed by the dideoxy chain-terminating method (34). Sequencing reactions were performed using α[^{32}P]-dATP (purchased from Dupont-New England Nuclear) and Klenow (purchased from Bethesda Research Labs) or α[^{35}S]-dATP(purchased from Dupont-New England Nuclear) and a Sequenase kit (obtained from United States Biochemicals). Each clone was sequenced 3-6 times. The reactions were displayed on 0.2mm X 30 cm X 60cm standard sequencing gels (35) which were doubly loaded over a period of five hours. Electrophoresis was performed at 90-95 watts (constant power). This resulted in a sequence pattern with at least a 50 base overlap between clones. Regions of sequence ambiguities which were noted using Klenow or GC compressions were resolved by using Sequenase and these gel electrophoresis conditions.

The sequence information was analyzed using programs accessed through the BIONET system (Intelligenetics, Inc.).

Mapping of the transcription start point of *fadR*.

The initiation site of transcription was evaluated by extension of a *fadR*-specific oligonucleotide hybridized to *in vivo* synthesized K12 mRNA using

reverse transcriptase. Total *in vivo* synthesized RNA was isolated from the *fadR*$^+$*E. coli* strain K12 grown in 100 ml LB broth to 0.35 OD_{550} using hot phenol (36). For mapping of the 5' end of *fadR*-specific RNA, the oligonucleotide 5'ATTCCAGATACTTTCAA 3' complementary to bases 83-99 in Figure 2 was 5'end-labelled using polynucleotide kinase and $\gamma[^{32}P]$-ATP (as detailed in reference 35) and then mixed with 20μg of RNA in 20 μl of buffer containing 250mM KCl and 10mM Tris, pH8.0. The samples were heated to 80^{o}C then rapidly transferred to a heating block set at the indicated hybridization temperature, usually 39^{o}C, for at least two hours. The hybridization temperature was estimated using the formula of Wallace and Miyada (37). After hybridization, the samples were diluted with 150μl of reverse transcriptase buffer containing 50mM Tris pH 7.3, 6mM $MgCl_2$, 40mM KCl, 2mM each dATP, dGTP, dCTP, dTTP and 28 units of AMV reverse transcriptase (purchased from United States Biochemicals, Inc.). Samples containing 20μg of total yeast tRNA or *fadR*$^+$ RNA and primer incubated without reverse transcriptase were used as controls. Enzymatic extension of the primer was allowed to proceed for 60min at 37^{o}C. Samples were precipitated with ethanol and resuspended in 10μl of formamide gel loading solution (34). The products of the primer extension reactions were displayed on a 60cm sequencing gel. Dideoxy sequencing reactions were performed on M13 subclone CD1901 using the same oligonucleotide as primer to determine the size of the extended fragment and position of the initiating nucleotide.

RESULTS AND DISCUSSION

Nucleotide sequence of the *fadR* gene.

In previous work the *fadR* gene was cloned on a 1.3kb *Hin*D III-*Eco*RV fragment (19). This fragment was determined to contain the entire *fadR* gene based on the following criteria: *fadR* strains harboring this clone 1) no longer were able to grow on medium chain fatty acids as a sole carbon and energy source; 2) had inducible levels of fatty acid oxidation and enoyl-CoA hydratase; and 3) had inducible levels of ß-galactosidase when the *fadR* strain also carried the operon fusion λΦ*fadE-lacZ*. Additionally, the cloned *fadR* gene relieved the requirement for unsaturated fatty acid supplementation of *fadR fabA*(Ts) strains at temperatures non-restrictive for the *fabA*(Ts) allele. In the present study, I have determined the nucleotide sequence of the *fadR* gene using the

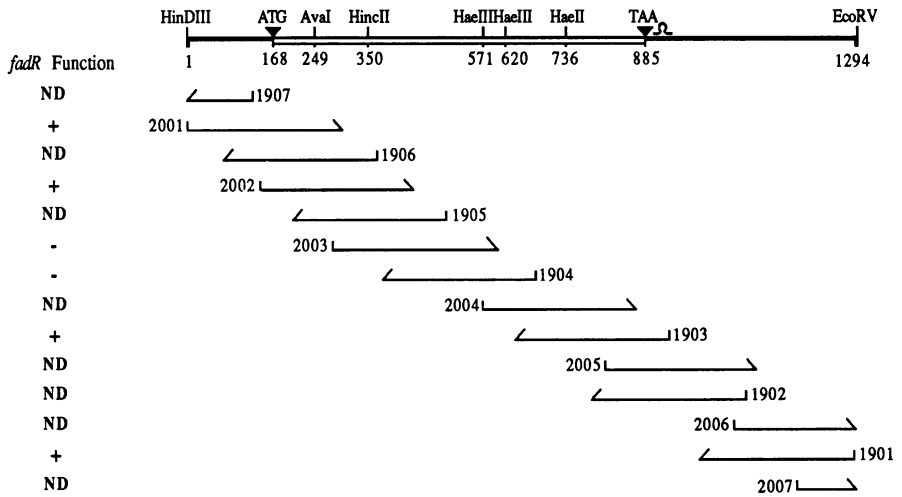

Figure 1. *fadR* sequencing strategy. The M13 clones for sequencing *fadR* were generated by cloning the 1.3kb DNA fragment encoding FadR into M13mp19 in both orientations. The series CD2001-CD2007 were sequenced from the *Hin*DIII site to the *Eco*RV site as illustrated. The series CD1901-CD1907 were sequenced from the *Eco*RV site to the *Hin*DIII site. The tail of the arrow indicates the position at which the sequencing initiates using the M13 universal primer. The tip of the arrow indicates the approximate distance each clone was read. The location of the initiation codon, the termination codon, and several restriction sites are also illustrated. To determine whether the deletions disrupted the coding sequence of FadR, subclones of *fadR* DNA were derived from the M13 replicative form, cloned into pUC18 or pUC19, and tested for their ability to confer *fadR*$^+$ phenotype, +, or not, -, as described in the text. ND, not determined.

dideoxy-chain terminating method. The *Hin*DIII-*Eco*RV fragment encoding *fadR* was cloned directly into M13mp19 in both orientations to generate CD1901 and CD2001. Each of these clones were used to generate sequential deletions as described above in Materials and Methods and illustrated in Figure 1. The complete nucleotide sequence of the 1.3 kb fragment encoding *fadR* is presented in Figure 2.

There was determined to be only one large open reading frame (ORF) within this 1294 base pair DNA fragment capable of encoding a polypeptide with a molecular weight of 26,972. This ORF extends from base pair 34 to base pair 750 (Figure2). There is only one methionine codon, ATG, at the 5'end of this ORF at base pairs 34-36. A sequence, 5'ATGGA 3', which correlates to the

```
 -130       -120       -110       -100       -90        80         -70
AAGCTTAACG GTCAGGCAGG AGTGAGGCAA GTCTTGATAG TCAAGGCGAA AGAGATGCGG AAAATGAAGC

  -60        -50        -40        -30        -20        -10        +1
CTTGATCCCT TTTTCTTCTT TTTGTCTGCT ATCAGCGTAG TTAGCCCTCT GGTATGATGA GTCCAACTTT
                                  -35                    -10        +1

  10         20         30          40          50          60
GTTTTGCTGT GTTATGGAAA TCTCACT ATG GTC ATT AAG GCG CAA AGC CCG GCG GGT TTC GCG
           ------          MET Val Ile Lys Ala Gln Ser Pro Ala Gly Phe Ala
            SD             ---

70          80          90          100         110         120
GAA GAG TAC ATT ATT GAA AGT ATC TGG AAT AAC CGC TTC CCT CCC GGG ACT ATT TTG
Glu Glu Tyr Ile Ile Glu Ser Ile Trp Asn Asn Arg Phe Pro Pro Gly Thr Ile Leu

  130         140         150         160         170         180
CCC GCA GAA CGT GAA CTT TCA GAA TTA ATT GGC GTA ACG CGT ACT ACG TTA CGT GAA
Pro Ala Glu Arg Glu Leu Ser Glu Leu Ile Gly Val Thr Arg Thr Thr Leu Arg Glu

      190         200         210         220         230         240
GTG TTA CAG CGT CTG GCA CGA GAT GGC TGG TTG ACC ATT CAA CAT GGC AAG CCG ACG
Val Leu Gln Arg Leu Ala Arg Asp Gly Trp Leu Thr Ile Gln His Gly Lys Pro Thr

          250         260         270         280         280
AAG GTG AAT AAT TTC TGG GAA ACT TCC GGT TTA AAT ATC CTT GAA ACA CTG GCG CGA
Lys Val Asn Asn Phe Trp Glu Thr Ser Gly Leu Asn Ile Leu Glu Thr Leu Ala Arg

300         310         320         330         340         350
CTG GAT CAC GAA AGT GTG CCG CAG CTT ATT GAT AAT TTG CTG TCG GTG CGT ACC AAT
Leu Asp His Glu Ser Val Pro Gln Leu Ile Asp Asn Leu Leu Ser Val Arg Thr Asn

    360         370         380         390         400         410
ATT TCC ACT ATT TTT ATT CGC ACC GCG TTT CGT CAG CAT CCC GAT AAA GCG CAG GAA
Ile Ser Thr Ile Phe Ile Arg Thr Ala Phe Arg Gln His Pro Asp Lys Ala Gln Glu

        420         430         440         450         460
GTG CTG GCT ACC GCT AAT GAA GTG GCC GAT CAC GCC GAT GCC TTT GCC GAG CTG GAT
Val Leu Ala Thr Ala Asn Glu Val Ala Asp His Ala Asp Ala Phe Ala Glu Leu Asp

470         480         490         500         510         520
TAC ACC ATA TTC CGC GGC CTG GCG TTT GCT TCC GGC AAC CCG ATT TAC GGT CTG ATT
Tyr Asn Ile Phe Arg Gly Leu Ala Phe Ala Ser Gly Asn Pro Ile Tyr Gly Leu Ile

   530         540         550         560         570         580
CTT AAC GGG ATG AAA GGG CTG TAT ACG CGT ATT GGT CGT CAC TAT TTC GCC AAT CCG
Leu Asn Gly MET Lys Gly Leu Tyr Thr Arg Ile Gly Arg His Tyr Phe Gla Asn Pro

       590         600         610         620         630
GAA GCG CGC AGT CTG GCG CTG GGC TTC TAC CAC AAA CTG TCG GCG TTG TGC AGT GAA
Glu ala Arg Ser Leu Ala Leu Gly Phe Tyr His Lys Leu Ser Ala Leu Cys Ser Glu

640         650         660         670         680         690
GGC GCG CAC GAT CAG GTG TAC GAA ACA GTG CGT CGC TAT GGG CAT GAG AGT GGC GAG
Gly Ala His Asp Gln Val Tyr Glu Thr Val Arg Arg Tyr Gly His Glu Ser Gly Glu

  700         710         720         730         740         750
ATT TGG CAC CGG ATG CAG AAA AAT CTG CCG GGT GAT TTA GCC ATT CAG GGG CGA TAA
Ile Trp His Arg MET Gln Lys Asn Leu Pro Gly Asp Leu Ala Ile Gln Gly Arg  •

    760         770         780         790         800         810         820
TCCCTTCCGT TTAAAGAGCA AACCCCTCAA ACGAGGGGTT TTTTGTTGTT TTTACAGATT TCCCATTCTT
                ------>           <------------- --

    830         840         850         860         870         880         890
GGCGGGCAAC GTTCCAGCAA CTCGATGCTG CCGTCTTCGT TTTGCTGTTC GAGCATCACA TCAAATCCCC

    900         910         920         930         940         950         960
ACAGGCGATG CACATGCTTC AGGACTTCTT TGCGCCCCCG ATCCCCTTCG AAGGAAAGAC CTGATGCTTT

    970         980         990        1000        1010        1020        1030
TCGTGCGCGC ATAAAATACC TTGATACTGT GCCGGATGAA AGCGGTTCGC GACGAGTAGA TGCAATTATG

   1040        1050        1060        1070        1080        1090        1100
GTTTCTCCGC CAAGAATCTC TTTGCATTTA TCAAGTGTTT CCTTCATGA TATTCCGAGA GCATCAATAT

   1110        1120        1130        1140        1150        1160
GCAATGCTGT TGGGATGGCA ATTTTTACGC CTGTTTTGCT TTGCTCGACA TAAAGCT
```

concensus sequence of Shine and Dalgarno (38) for the ribosome binding site occurs closely upstream. The deduced *fadR* coding sequence ends with a termination codon TAA at base pairs 751-753. This region is followed by a GC-rich region of dyad symmetry (estimated hairpin formation $\Delta G = -15.3$) which may be sufficient for factor independent termination (39). No other regions of extensive dyad symmetry predicted to form stable stem and loop structures were found.

Previous work employing maxicell analysis had identified a single polypeptide associated with wild-type *fadR* gene activity with an apparent molecular weight of 29,000 as determined by polyacrylamide-sodium dodecyl sulfate gel electrophoresis (19). This estimation of molecular weight is in reasonable agreement with the molecular weight predicted from the DNA sequence to encode the *fadR* protein (26,954). This ORF encompasses the *Hinc*II (350bp) and the *Hae*II (736bp) restriction endonuclease cleavage sites previously shown by subcloning experiments and maxicell analysis to disrupt *fadR* gene activity (19). To more narrowly define the *fadR* coding sequence and to lend support to the conclusion that this was the correct reading frame, I subcloned several *Exo*III deletions of CD2001 and CD1901 by restriction of the M13 clone of interest with *Hin*DIII and *Eco*RI (See Figure 1). These restriction enzyme cleavage sites lie outside of the insert of *E. coli* chromosomal DNA in the polylinker region of M13mp19. The fragments containing *fadR* sequences of interest were inserted into pUC18 or pUC19 such that the predicted *fadR* coding sequence was in opposite orientation to the strong *lac* promoter. The resulting plasmids incude: pCD110 a subclone of CD2001 containing the entire 1294 base pair sequence; pCD111 a subclone of CD2002 in which bases -134 to -24 are deleted; pCD112 a subclone of CD1903 in which bases +806 to +1160 are deleted; and pCD114 a subclone of CD2003 in which bases -134 to +106 are deleted.

The presence of *fadR*$^+$ was evaluated by transforming the plasmid of interest into *fadR*::Tn*10* strain LS1085. Transformants were selected on LB

Figure 2. The complete nucleotide sequence and derived amino acid sequence of *fadR*. The transcription initiation site is labeled +1 and is delineated by a vertical arrow. The series of nucleotides closest to concensus for *E. coli* promoters are underlined and labeled -10 and -35. The putative ribosome binding site is underlined and labeled SD. The region of dyad symmetry proposed to be the factor independent transcription terminator is marked by two convergent arrows.

plates containing 100μg/ml ampicillin. The *fadR* phenotype of the transformants was evaluated by replica plating to minimal media containing 5mM oleate ($C_{18:1}$) and ampicillin or 5mM decanoate (C_{10}) and ampicillin. In this case, cells carrying intact *fadR* on a plasmid will be able to grow on minimal media containing oleate but not minimal media containing decanoate. This is because long- but not medium-chain fatty acids can induce the expression of the *fad* enzymes. Inactivation of the negative regulatory gene, *fadR*, is required for growth on medium-chain fatty acids. As an additional test for the presence of *fadR*$^+$, the plasmids were transformed into strain LS1155, *fadR*::Tn10/ λΦ*fadA-lacZ*, and the transformants were plated on MacConkey indicator plates containing 1% lactose and 100μg/ml ampicillin. In the *fadA-lacZ* fusion, the gene for ß-galactosidase is under the control of the *fadA* promoter. Since *fadR* is inactivated in strain LS1155 due to the presence of the transposon, ß-galactosidase activity is constitutive and the strain was red on the indicator plates. Transformants harboring plasmids which carry *fadR*$^+$ were white on these indicator plates while transformants harboring plasmids with *fadR* deletions or the vector alone were red. Based on these two independent methods used to characterize the *fadR* phenotyes, each of the plasmids constructed above except pCD114 carried an intact *fadR* gene. This is consistent with the predicted ORF. The only result which was somewhat surprising was that pCD111 in which 110 bases of the 5' sequences are deleted also appears *fadR*$^+$. This deletion therefore extends into the predicted -35 region of the *fadR* promoter as discussed below. This result may be explained if the deletion generates a sequence which may act as a promoter for *fadR*. Sequencing of CD2002 using a primer complementary to sequences upstream of the *Eco*RI site yields the following sequence: 5'-TTGTAAAACGA-N_{17}-TATGAT-3'. This construction maintains the predicted -10 region and yields a poor -35 region. Since the plasmid pCD111 is present in very high copy number within the cell, this construction may be sufficiently transcribed to give a *fadR*$^+$ phenotype. Deletions into the predicted coding region as in clones pCD114 described in this work and pACHinc1 and pACHH8 described previously (19) do not confer the wild-type *fadR* phenotype even in the multicopy condition.

Mapping of the 5'end of *fadR* mRNA.

The initiation site of transcription was determined using primer extension analysis. These experiments are described in Materials and Methods. A repre-

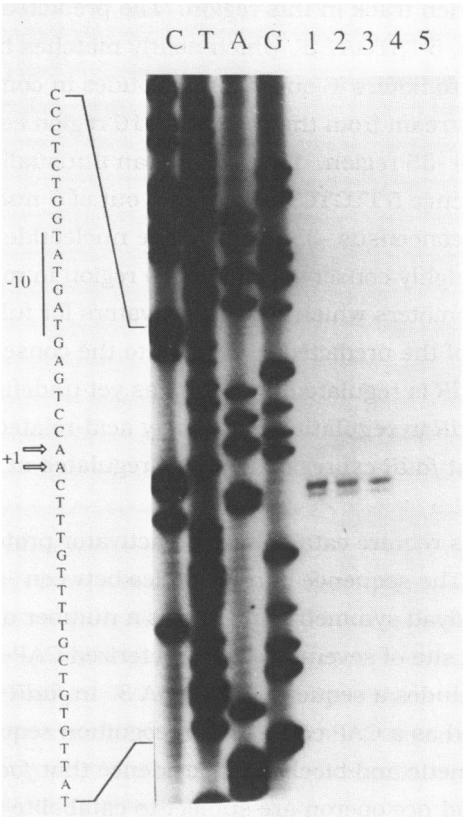

Figure 3. Determination of the site of *fadR* transcription initiation by primer extension. An *fadR*-specific oligonucleotide was 5' end labeled with ^{32}P, hybridized to RNA isolated from *E. coli* K12, and extended with reverse transcriptase as described within the text. Lanes C, T, A and G are the products of dideoxy sequencing reactions (G, A, T and C respectively) using the same oligonucleotide as primer and CD1901 as template. In this experiment samples were hybridized at 39°C (lanes 1, 4 and 5), 42°C (lane 2) or 46°C (lane 3). Lane 4 is the product of the hybridization reaction without extension (*i.e.* no reverse transcriptase). Lane 5 is the product of hybridization and extension of yeast tRNA using the same *fadR*-specific primer.

sentative experiment is presented in Figure 3. The most prominent extended fragment indicates that intitiation occurs with an adenine residue, designated +1. The primer extension experiments also yielded a less pronounced band at nucleotide +11 which we believe is due to a failure of the enzymatic extension

through the uridine rich track in this region. The predicted *fadR* promoter includes a -10 region, 5'TATGAT 3', which nearly matches the concensus sequence for *E. coli* promoters (6 out of 7 nucleotides in common). The region 15-19 base pairs upstream from the predicted -10 region corresponds less well to the consensus -35 region. If one allows an unusually wide spacing of 25 nucleotides the sequence 5'TTGTCT'3 shares 4 out of 6 nucleotide identities in common with the concensus -35 region. The nucleotide sequence in the -35 region is not as highly conserved as the -10 region in many *E. coli* genes (40), particularily promoters which require activators for full expression (41). The poor homology of the predicted -35 region to the consensus sequence may indicate that *fadR* is regulated by factors as yet undefined. Given the important role for *fadR* in regulating many fatty acid-related pathways it would not be surprising that *fadR* expression is itself regulated at the level of transcription.

Many *E. coli* genes require catabolite gene activator protein (CAP) and cAMP for full expression. The sequence of nucleotides between -114 and -90 in Figure 2 has partial dyad symmetry and shows a number of nucleotide identities with the binding site of several well-characterized CAP-responsive promoters. This includes a sequence 5'AGTGA 3' in *fadR* which is similar to 5'TGTGA 3' suggested as a CAP concensus recognition sequence (42). While there is extensive genetic and biochemical evidence that *fadR*-reponsive genes of the *fad* regulon and *ace* operon are subject to catabolite repression and require cAMP-CAP for full expression this is the first indication that *fadR* might be subject to similar controls.

Predictions of protein structure from the derived amino acid sequence.

A single large ORF of 239 amino acids is predicted to encode the *fadR* protein. The derived amino acid sequence is given in Figure 2. The codon usage for this polypeptide is similar to the codon usage for moderately expressed genes (44) including many *E. coli* regulatory genes (45). There is no unusual clustering or extensive use of rare codons.

The *fadR* gene product has been previously characterized by classical genetic means as a diffusible protein which exerts negative control over fatty acid degradation and acetate metabolism by decreasing the transcription of the *fad* and *ace* structural genes (1). Many *E. coli* regulatory proteins which exert their affects by binding to DNA share amino acid and structural identities over at least a 20 amino acid segment generally found near the amino

terminus, termed the helix-turn- helix motif (46-48). Amino acids 34-54 of the *fadR* peptide sequence presented in Figure 2 shows reasonable homology to the helix-turn-helix motif. This segment is : glu_{34}-arg-glu-leu-ser-glu-leu-ile-gly-val-thr-arg-thr-thr-leu-arg-glu-val-leu-gln_{53}. Structural predictions of this *fadR* peptide segments using the method of Chou and Fasman(49) predict α-helical segments for amino acids 34-41 (P_α=1.145; P_β=0.958) and amino acids 46-53 (P_α=1.111; P_β=1.003) and a ß-sheet segment for amino acids 40-47 (P_α=0.924; P_β=1.231). Within the ß-sheet segment is found the highly conserved glycine residue. Residues 37, 41,43,and 48 are hydrophobic while residues 34, 35, 36 39, 44, 45, and 47 are hydrophilic in agreement with the predictions of the model of Pabo and Sauer (46). Analysis of this segment by the method of Dodd and Egan (50) which estimates the occurence of each amino acid at each position relative to a master set of DNA-binding domains yields a low (6%) but significant probability that this peptide segment resembles the λCro DNA-binding domain. While this peptide sequence comparison only suggests a DNA-binding region within the *fadR* protein, it does provide a basis on which to design future studies to evaluate whether or not this segment of the *fadR* protein in particular is important in regulation of *fadR*-responsive genes and if this regulation is mediated by DNA-binding.

The method of Kyte and Doolittle (51) for the analysis of the hydropathic character of a protein was applied to the ORF predicted to encode *fadR*. This analysis predicts that *fadR* is a soluble protein. There is only one region with non-polar character which might predict a membrane spanning region for this protein (*i. e.* 20 amino acids or more (52)). It is unclear if this structural prediction reflects actual membrane association.

The DNA and derived amino acid sequences presented in this study will be used to design future experiments directed toward a detailed characterization of the *fadR* gene and protein. The ultimate goal of this work is to describe in molecular detail the control of the *fad, ace* and *fab* genes by *fadR*. This regulatory gene modulates these genetically distinct, biochemically interrelated pathways in a complex manner which is presently poorly understood.

ACKNOWLEDGEMENT

The oligonucleotides used in the dideoxy sequencing and primer extension experiments were provided by the Molecular Resource Center at the University

of Tennessee, Memphis. The BIONET computer network is supported by a grant awarded by the NIH Division of Research Resources, number RR01865-05.

This work was supported by a grant from the University Physicians Foundation, Inc., Memphis, Tn.

REFERENCES

1. Nunn, W. D. (1986) Microbiol. Rev. **50**, 179-192.
2. Nunn, W.D. and Simons, R. W. (1978) Proc. Natl. Soc. USA **75**, 3377-3381.
3. Ginsburgh, C. L., Black, P.N., and Nunn, W.D. (1984) J. Biol. Chem. **260**, 8437-8443.
4. Black, P. N., Kianian, S. F., Di Russo, C. C., and Nunn, W. D. (1985) J. Biol. Chem. **260**, 1780-1790.
5. Black, P. N., Said, B., Goshn, C., Beach, J. V. and Nunn, W. D. (1987) J. Biol. Chem. **262**, 1412-1419.
6. Overath, P., Pauli, G., and Shairer, H. U. (1969) Eur. J. Biochem. **7**, 559-574.
7. Frerman, F. E. and Bennett, W. (1973) Arch. Biochem. Biophys. **159**, 434-443.
8. Rock, C. O. and Cronan, J. E., Jr. (1985) In Vance, D. E. and Vance, J. E. (eds.), Biochemistry of Lipids and Membranes, Benjamin/Cummings Publishing Co., Inc. Menlo Park, Ca., pp.73-115.
9. Overath, P., Raufuss, E., Stoffel, W. and Ecker, W. (1967) Biochem. Biophys. Res. Comm., **29**, 28-33.
10. Weeks, G., Shapiro, M., Burns, R.O. and Wakil, S.J. (1969) J. Bacteriol. **97**, 827-836.
11. Klein, K., Steinberg, R., Fiethen, B., and Overath, P. (1971) Biochemistry **19**, 442-450.
12. Kornberg, H. L., (1966) Biochem. J. **99**, 1-11.
13. Brice, C. G. and Kornberg, H. L. (1968), J. Bacteriol. **96**, 2185-2186.
14. Lakshmi, T.M. and Helling, R. B. (1978) Can. J. Microbiol. **24**, 149-153.
15. Gottseman, S. (1984) Ann. Rev. Genet. **18**, 415-441.
16. Simons, R.W., Hughes, K.T., and Nunn, W.D. (1980) J. Bacteriol. **142**, 621-632.
17. Simons, R.W., Egan, P.A., Chute, H.T. and Nunn, W.D., (1980) J. Bacteriol. **143**, 726-730.
18. Maloy, S.R. and W.D. Nunn, (1982) J. Bacteriol. **149**, 173-180.
19. Di Russo, C.C. and Nunn, W.D. (1985) J. Bacteriol. **161**, 583-588.
20. Sallus, L., Haselbeck, R.J., and Nunn, W.D., (1983), J. Bacteriol. 155: 1450-1454.
21. Clark, D. , (1981), J. Bacteriol. **148**, 521-526.
22. Maloy, S.R., Bohlander, M. and Nunn, W.D. (1981), J. Bacteriol. **143**, 720-725.
23. Vanderwinkel, E., De Vliegher,M., Fontaine, M., Charles, D., Denamur, F., Vandervoorde, D., De Kegel, D. (1976) J. Bacteriol. **127**, 1389-1399.
24. Donachie, W.D., Begg, K.J., and Sullivan, (1984), In Losick, R. and Shapiro, L. (eds), Microbial Development, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., pp.27-62.
25. Cronan, J. E. ,Jr and Gelman, E.P. (1973) J. Biol. Chem. **248**, 1188-1195.

26. Cronan, J.E., Silbert, D.F., and Wulff, D.L. (1972) J. Bacteriol. **112**, 206-211.
27. Nunn, W.D., Giffen, P.K., Clarke, D., and Cronan, J.E.,Jr. (1981) J. Bacteriol. **134**, 554-560.
28. Messing,J. (1983) Methods in Enzymol. **101**, 20-78.
29. Daghert, M. and Ehrlich, S.D. (1979) Gene **6**, 23-28.
30. Miller, J.H. (1972) Experiments in Molecular Genetics, Cold Spring Harbor Laboratories, Cold Spring Harbor, N.Y.
31. Volgel, J.H. and Bonner, D.M. (1956) J. Biol. Chem. **218**, 97-106.
32. Komeda, Y. and Ino, T. (1979) J. Bacteriol. **139**, 721-729.
33. Henikoff, S. (1984) Gene **28**, 351-359.
34. Sanger, F., Nicklen,S., and Coulsen, A.R. (1977) Proc. Natl. Acad. Sci. USA **74**, 5463-5467.
35. Maxam, A.M. and Gilbert, W. (1980) Methods in Enzymol. **65**, 499-560.
36. Markam. B.E., Harper, J.E., Mount, D.W., Sancar, G.B., Sancar, A., Rupp, W.D., Kenyon, C.J. and Walker, G.C. (1984) J. Mol. Biol. **178**, 237-248.
37. Wallace, R.B. and Miyada, C.G. (1987) Methods in Enzymol. **152**, 432-442.
38. Shine. J. and Dalgarno, L. (1975) Nature **254**, 34-38.
39. Platt, T. (1986) Ann. Rev. Biochem. **55**, 339-372.
40. Hawley, D.K. and McClure, W.R. (1983) Nucl. Acids Res. **11**, 2237-2255.
41. Raibaud, O. and Schwartz, M. (1984) Ann. Rev. Genet. **18**, 173-206.
42. deCrombrugghe, B., Busby, S., and Buc, H. (1984) Science **224**, 831-838.
43. Kolb, A., Spassky, A., Chapon, C., Blazy, B. and Buc, H. (1983) Nucl. Acids Res. **11**, 7833-7838.
44. Ikemura, T. (1981) J. Mol. Biol. **151**, 389-409.
45. Sharp, P.M. and Li, W-H. (1986) Nucl. Acids Res. **14**, 7737-749.
46. Pabo, C.O. and Sauer, R.T. (1984) Ann. Rev. Biochem. **53**, 293-321.
47. Matthews, B.W., Ohlendorf, D.H., Andersen, W.F. and Takeda, Y. (1982) Proc. Natl. Acad. Sci. USA **79**, 1428-1432.
48. Sauer, R.T., Yocum, R.R., Doolittle, R.F., Lewis, M. and Pabo, C.O. (1982) Nature **298**, 447-451.
49. Chou, P.Y. and G.D. Fasman. (1978) Ann. Rev. Biochem. **47**, 251-276.
50. Dodd, I.B. and J.B. Egan. (1987) J. Mol. Biol. **194**, 557-564.
51. Kyte, J. and Doolitttle, R.F. (1982) J. Mol. Biol. **157**, 105-132.
52. Engelman, D.M., Steitz, T.A., and Goldman, A. (1986) Ann. Rev. Biophys. Biophys. Chem. **15**, 321-353.