JOURNAL OF BACTERIOLOGY, Feb. 1989, p. 791–798
0021-9193/89/020791-08$02.00/0

Vol. 171, No. 2

Nucleotide Sequence of a Cluster of *Escherichia coli* Enterobactin Biosynthesis Genes: Identification of *entA* and Purification of Its Product 2,3-Dihydro-2,3-Dihydroxybenzoate Dehydrogenase

JUN LIU, KENNETH DUNCAN,† AND CHRISTOPHER T. WALSH*

Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, 25 Shattuck Street, Boston, Massachusetts 02115

Received 12 August 1988/Accepted 26 October 1988

The nucleotide sequence of a region of the *Escherichia coli* chromosome encoding part of a cluster of genes involved in the biosynthesis of the iron chelator enterobactin has been determined. Four closely linked open reading frames, corresponding to the coding regions of *entE* (carboxy-terminal 144 amino acids), *entB* (32,554 daltons), *entA* (26,249 daltons), and an unidentified gene (*P15*) encoding a 14,970-dalton protein, were found. The lack of intergenic sequences and promoterlike elements suggests that these genes form part of the same transcription unit. We report the purification to homogeneity of the *entA* product, 2,3-dihydro-2,3-dihydroxybenzoate dehydrogenase. It is an octamer of native molecular weight 210,000; the amino-terminal amino acid sequence confirmed the *entA* coding region. No isochorismate synthase activity was associated with this polypeptide. This finding leads to the conclusion that the recent suggestion (M. S. Nahlik, T. P. Fleming, and M. A. McIntosh, J. Bacteriol. 169:4163–4170, 1987) that 2,3-dihydro-2,3-dihydroxybenzoate dehydrogenase and isochorismate synthase activities reside on a single 26,000-dalton bifunctional enzyme is incorrect, even though the *entA* and *entC* mutations map to the same genetic locus.

Escherichia coli is able to sequester iron from its environment by synthesizing an iron-chelating compound, or siderophore, called enterobactin (enterochelin). Enterobactin is a cyclic triester of 2,3-dihydroxybenzoyl serine which is secreted into the growth medium, whereupon it is able to remove iron bound to transferrin or lactoferrin in serum and body fluids; specific transport systems then scavenge the Fe^{3+}-enterobactin complexes and release the iron into the cell cytoplasm, where it is available for cellular metabolism (for reviews, see references 2 and 7).

The biosynthesis of enterobactin begins with chorismic acid, the end product of the shikimate, or common aromatic amino acid, pathway. The first three steps of the biosynthesis have been characterized biochemically in crude cytoplasmic extracts of *Enterobacter aerogenes* (33–35). Chorismate is isomerized to isochorismate by isochorismate synthase (the *entC* gene product); the enolpyruvyl side chain of isochorismate is removed by the *entB* gene product (2,3-dihydro-2,3-dihydroxybenzoate [diDHB] synthetase), yielding diDHB, which is then oxidized to 2,3-dihydroxybenzoate by diDHB dehydrogenase (diDHB-DH, the *entA* gene product). In a less well characterized process, L-serine and 2,3-dihydroxybenzoate are activated and linked to form enterobactin by enterobactin synthetase, which is believed to be a membrane-bound multienzyme complex consisting of the products of the *entD*, *entE*, *entF*, and *entG* genes (11, 12, 32).

The enterobactin biosynthetic and transport genes are clustered on the *E. coli* chromosome around min 13 (36). Clones that cover overlapping regions of the cluster have been obtained in a number of laboratories. Somewhat conflicting evidence was obtained regarding the order of the genes involved in synthesis of the intermediate 2,3-dihydroxybenzoate (*entA*, *entB*, and *entC*) and the number and nature of transcription units (8, 9, 15, 23). Nahlik et al. (19) recently defined the gene order as *entEBG(AC)* by deletion analysis, correlated with gene expression in *E. coli* minicells, and suggested that the *entC* and *entA* activities were expressed on a single 26,000-dalton bifunctional polypeptide.

In this paper, we report the DNA sequence of a 3.25-kilobase (kb) fragment of *E. coli* DNA within the *ent* gene cluster, which encodes the carboxy-terminus of *entE*, the entire *entB* and *entA* coding regions, and the sequence of an as yet unidentified open reading frame (ORF) that encodes a 14,970-dalton polypeptide. We have constructed an overproducing *E. coli* strain that has the *entA* structural gene under the control of the *tac* promoter. This strain produces diDHB-DH as greater than 6% of total cell protein. We also report the first purification to homogeneity of an *E. coli* enterobactin biosynthetic enzyme, diDHB-DH, and show that isochorismate synthase activity is not associated with this 26,249-dalton polypeptide.

MATERIALS AND METHODS

Materials. L-[^{35}S]methionine and [α-^{35}S]thio-dATP were from Amersham Corp., Arlington Heights, Ill. Isopropyl-β-D-thiogalactoside (IPTG) was from Sigma Chemical Co., St. Louis, Mo. Cyclone I kit was from International Biotechnologies Inc., New Haven, Conn. Modified bacteriophage T7 DNA polymerase was a gift from Stan Tabor of this department. DiDHB (racemic form) was from Nina Quinn and Glenn Berchtold (Chemistry Department, Massachusetts Institute of Technology, Cambridge). 2,3-Dihydroxybenzoate was purchased from Aldrich Chemical Co., Milwaukee, Wis. Chorismic acid was prepared by the procedure of Gibson (10).

Bacterial strains and plasmids. *E. coli* JM101 [*supE thi* Δ(*lac-proAB*) (F′ *traD36 proA*$^+$ *proB*$^+$ *lacI*^{q} *lacZ*ΔM15)] and JM105 [Δ(*lac-pro*)*XIII thi rpsI* (Str^r) *endA sbcB supE hsdR* F′ *traD36 proA*$^+$ *proB*$^+$ *laqI*^{q} *lacZ*ΔM15], pUC18, pUC19,

* Corresponding author.
† Present address: Glaxo Group Research Ltd., Greenford, Middlesex UB6 0HE, United Kingdom.

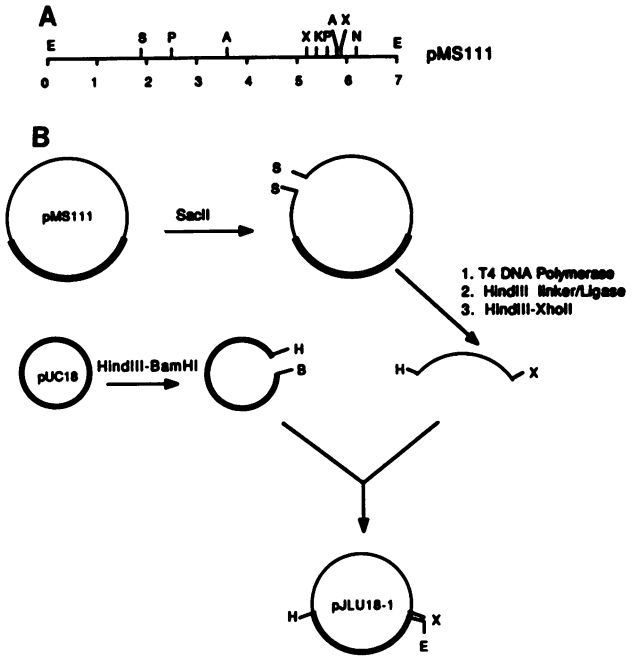

FIG. 1. Restriction map of pMS111 and subcloning of a 3.25-kb *Sac*II-*Xho*II fragment into pUC18. For the subcloning scheme, the heavier line represents vector. Abbreviations: A, *Acc*I; B, *Bam*HI; E, *Eco*RI; EV, *Eco*RV; H, *Hin*dIII; K, *Kpn*I; N, *Nde*I; P, *Pvu*II; S, *Sac*II; X, *Xho*II.

and the replicative forms of M13mp18 and M13mp19 were obtained from New England BioLabs, Beverly, Mass. pKK223-3 was from Pharmacia Fine Chemicals, Piscataway, N.J. *E. coli* K38(pGP1-2) and pT7-7 were obtained from Stan Tabor. pMS111, *E. coli* AN194 (F^- *pro leu trp* Str^r), AN91 (*proA2 argE3 pheA1 tyrA4 trp401 gal401* Str^r *entA404*), AN191 (*entC401* derivative of AN194), and AN192 (*entB402* derivative of AN194) were provided by I. G. Young, Australian National University.

DNA manipulations. DNA manipulations were carried out as described previously (18). Enzymes were obtained from a number of suppliers and were used in accordance with the instructions of the manufacturers.

DNA sequencing. A contiguous series of clones for sequencing was generated by sequential deletion of a linearized single-stranded DNA clone by treatment with T4 DNA polymerase (6). The dideoxynucleotide termination method of Sanger et al. (25) was used to obtain the DNA sequences, using a modified form of T7 DNA polymerase as previously described (29). Sequence ambiguities were resolved by using dITP in place of dGTP in some reactions. The sequence data obtained were analyzed by the ANALYSEQ program of Staden (26).

Pulse labeling of plasmid-encoded polypeptides. *E. coli* K38(pGP1-2), which has the T7 RNA polymerase gene under the control of bacteriophage λ p_L (cI857), was transformed with pJLT7-1A, pJLT7-1B, pJLT7-2A, and pJLT7-2B. The proteins encoded by these plasmids were labeled with L-[^{35}S]methionine as described by Tabor and Richardson (28).

Enzyme assays. (i) diDHB-DH. It was found that the product of the reaction catalyzed by diDHB, 2,3-dihydroxybenzoate, has a fluorescence emission at 437 nm when excited at 306 nm. The fluorescence absorption at 306 (excitation)/437 (emission) was therefore used to determine diDHB-DH activity. The amount of product released was

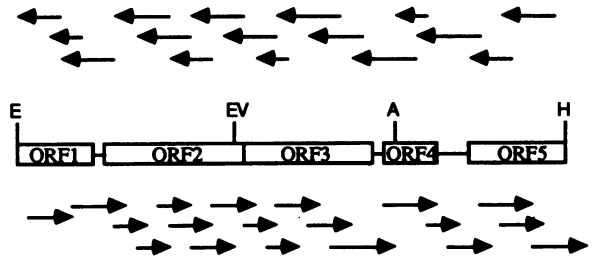

FIG. 2. Sequencing strategy. Arrows represent the direction and length of sequenced clones.

quantified by correlation with a standard curve made from pure 2,3-dihydroxybenzoate. One unit of activity is defined as the quantity of enzyme required to convert 1 μmol of substrate to product min^{-1} at 37°C. The assay mixture contained 90 mM potassium phosphate (pH 7.2), 2 mM NAD^+, and 1 mM diDHB in a final volume of 1 ml. An LS-3 fluorescence spectrometer (The Perkin-Elmer Corp., Norwalk, Conn.) linked to a constant-temperature water bath at 37°C was used throughout.

(ii) Isochorismate synthase. Enzyme activity was assayed by monitoring the conversion of chorismate (λ_{max} = 273 nm; ε = 2,630 M^{-1} cm^{-1}) to isochorismate (λ_{max} = 278 nm; ε = 13,000 M^{-1} cm^{-1}) at 275 nm. Assays (1 ml) contained 100 mM Tris buffer (pH 8.0), 5 mM $MgCl_2$, and 40 mM chorismate.

(iii) Protein. Protein concentrations were determined by the method of Bradford (3).

Preparation of diDHB-DH. (i) Crude cell extract. A 1-liter culture of *E. coli* JM105(pJLK-2) was grown in L broth containing ampicillin (100 μg ml^{-1}) until the A_{595} reached 0.7; IPTG was added to a final concentration of 2 mM, and incubation continued for an additional 6 h. Cells were harvested, washed in 10 mM potassium phosphate buffer (pH 7.4) (KP_i buffer), and resuspended in 10 ml of the same buffer. Cells were lysed by two passages through a French press operating at 12,000 lb in^{-2} at the orifice. Cell debris was removed by centrifugation (30 min, 10,000 × *g*), and nucleic acids were precipitated by the addition of one-fifth volume of 2% (wt/vol) protamine sulfate to yield a crude extract.

(ii) Ion-exchange chromatography. The crude cell extract was applied to a 15-by-1.5-cm column of DEAE-Sepharose CL-6B (Pharmacia) and washed with KP_i buffer until the A_{280} was <0.1. The column was eluted with a 0 to 0.5 M KCl gradient in KP_i buffer at a flow rate of 20 ml h^{-1}. Activity-containing fractions were pooled and dialyzed against KP_i buffer.

(iii) Dye ligand chromatography. The pooled, dialyzed material from step ii was applied to a column of Dyematrex Blue A agarose (7.0 by 1.0 cm; Amicon Corp., Lexington, Mass.) in KP_i. After extensive washing with the same buffer, the enzyme was eluted with a gradient of 0 to 1 M KCl in KP_i. Active fractions were pooled and concentrated by ultrafiltration. This material was dialyzed against KP_i buffer containing 50% (vol/vol) glycerol for long-term storage at −20°C.

Molecular weight determination. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis was performed as described previously (14). The molecular weight of native enzyme was determined by chromatography on an FPLC Superose 12 column (Pharmacia), using the following proteins as standards: thyroglobulin (669,000 daltons), ferritin

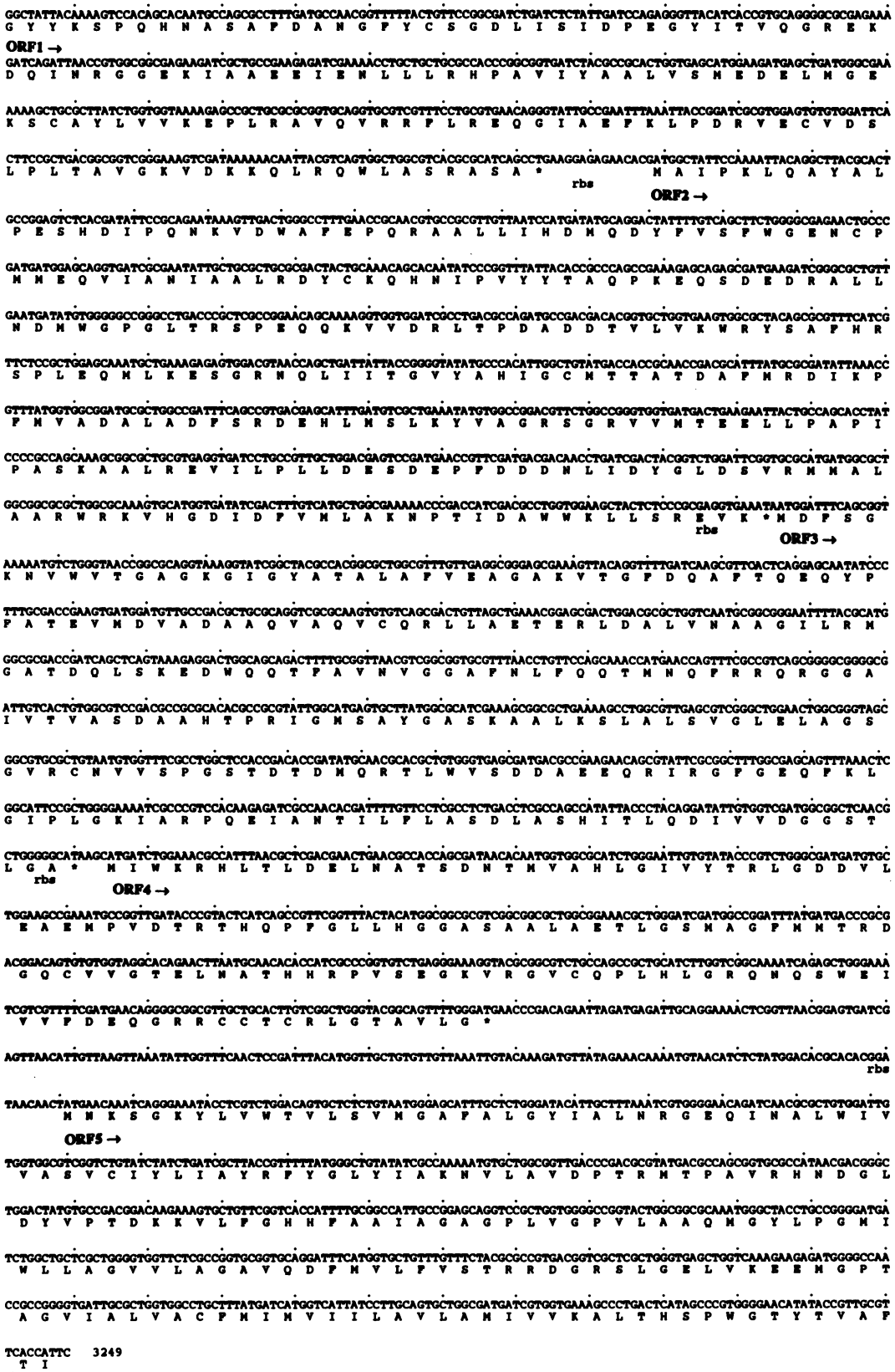

FIG. 3. Nucleotide sequence of the 3.25-kb fragment together with the amino acid translation of the longest ORFs. The positions of ORF1, ORF2, ORF3, ORF4, ORF5, and potential ribosome-binding sites (rbs) are indicated below the sequences.

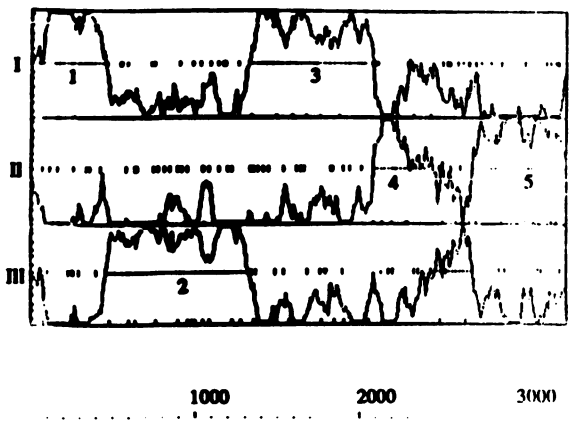

FIG. 4. Identification of likely gene coding regions by codon preference with the *E. coli* coding probability plots (window = 25 codons) of the ORFs of the 3.25-kb DNA sequence. Dotted lines across the half-point of each frame correspond to 50% coding probability. Vertical dashes across the half-points represent stop codons, and those at the bottom of each box represent ATG codons.

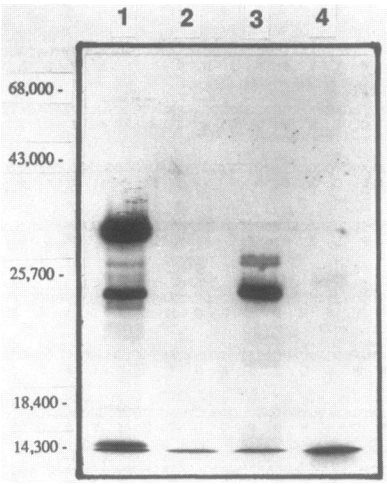

FIG. 5. Pulse-labeling of gene products from plasmids. Lanes: 1, pJLT7-1A; 2, pJLT7-1B; 3, pJLT7-2A; 4, pJLT7-2B. Molecular weight standards were lysozyme (14,300); β-lactoglobulin (18,400); α-chymotrypsinogen (25,700); ovalbumin (43,000); bovine serum albumin (68,000); and phosphorylase (97,400).

(440,000 daltons), catalase (232,000 daltons), aldolase (158,000 daltons), and bovine serum albumin (68,000 daltons).

RESULTS

Restriction mapping of pMS111 and isolation of *entA*. From previously reported complementation tests (15), pMS111 was believed to encode the *entA*, *entB*, *entC*, and *entE* genes. To further localize these genes, pMS111 was mapped with a number of restriction enzymes to identify new and useful restriction sites (Fig. 1A). A 3.25-kb *Sac*II-*Xho*II fragment was subcloned (Fig. 1). pMS111 was linearized with *Sac*II, the ends were repaired by treatment with T4 DNA polymerase in the presence of all four deoxynucleotide triphosphates, and *Hin*dIII linkers were added. After digestion with *Hin*dIII and *Xho*II, a 3.25-kb fragment was isolated and ligated into pUC18 (digested with *Hin*dIII and *Bam*HI) to give pJLU18-1. pJLU18-1 was transformed into *E. coli* mutants AN91, AN191, and AN192, and the phenotypes of transformants were examined. pJLU18-1 was found to complement the *entC*, *entA*, and *entB* mutations in these strains.

DNA sequence determination. The deletion method of Dale et al. (6) was used to generate a full set of overlapping clones in both directions of the 3.25-kb *Sac*II-*Xho*II fragment (as cloned in pJLU18-1) for sequence analysis. This fragment was cloned into M13mp19 and M13mp18, and single-stranded DNA was isolated. Deletion clones were made from insert cloned in M13mp19 by the procedure described in the Cyclone I kit. Difficulties encountered in digesting to completion the single-stranded DNA cloned into M13mp18 necessitated the use of a different strategy for the production of clones on the opposite strand. The *Eco*RI site adjacent to the insert in pJLU18-1 was modified to a *Hin*dIII site after treatment with the Klenow fragment of *E. coli* DNA polymerase I in the presence of deoxynucleotide triphosphates, followed by addition of *Hin*dIII linkers. The resulting 3.25-kb *Hin*dIII-*Hin*dIII fragment was cloned into M13mp19, and clones inserted into the appropriate orientation were selected. Single-stranded DNA was prepared and treated as described above to yield a full set of deletion clones.

The nucleotide sequences of these clones were determined and compiled (Fig. 2). Both strands of the 3.25-kb fragment were sequenced by using overlapping deletion clones. The sequence is shown in Fig. 3, together with an amino acid translation showing only long ORFs. The fragment contained three complete ORFs and parts of two others. The DNA sequence was analyzed by using an ANALYSEQ routine which predicts the likelihood that an ORF codes for a protein by measuring the distribution of codons over the sequence and comparing this with an average of preferred codon usage within other *E. coli* genes (26, 27). A plot of predicted coding regions (Fig. 4) revealed a high probability that all five ORFs were expressed genes. The complete sequences (ORF2, ORF3, and ORF4) were found to encode polypeptides of 285, 248, and 137 amino acids, with calculated molecular weights of 32,554, 26,249, and 14,970, respectively, from the first ATG-Met codon. ORF1 encodes the carboxy-terminal 144 amino acids of a protein, and ORF5 encodes the first 200 amino acids of another protein. By correlation with complementation and in vivo transcription-translation data (19), we propose that ORF1 corresponds to *entE*, ORF2 is *entB*, and ORF3 is *entA*. ORF4 has been designated *P15*. A notable feature of the sequence is the almost complete lack of intergenic sequences separating the coding regions: *entE* and *entB* are separated by 13 nucleotides, *entB* and *entA* overlap by 1 nucleotide, and the *entA-P15* boundary is 2 nucleotides. The identity and function of ORF5 are unknown.

The amino acid sequences predicted by the *entB*, *entA*, and *P15* genes were compared with the Protein Identification Resource protein sequence data base of known protein sequences, using the program FASTP (16). No proteins with significant sequence identity to *entB* and *P15* products were found, but a low-level sequence identity was found between the *entA* product and the ribitol dehydrogenase from *Klebsiella aerogenes* (17), which suggested that *entA* encodes the NAD-utilizing diDHB-DH. A comparison was also made by scanning the sequence of ORF3 for the typical βαβ-fold fingerprint described elsewhere (31). Between residues 7 and 36 of ORF3, 8 of the 11 features of the fingerprint were found, which provided a clear indication that this is an NAD-binding domain. If *entA* encodes a bifunctional

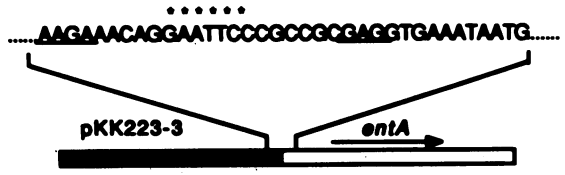

FIG. 6. Partial sequence of the *entA* product overproducer pJLK-2. The *Eco*RI site bordering the pKK223-3 vector and the *entA* gene is indicated by asterisks. Two ribosome-binding sites, one from the vector (left) and the other from the *entA* gene (right), are underlined. Sequences from the *entA* gene include the starting ATG codon.

diDHB-DH–isochorismate synthase, then one might have expected to detect sequence identity between this enzyme and the products of the *E. coli trpE* (anthranilate synthase) and *pabB* (*para*-aminobenzoate synthase) genes, since all three use chorismate as substrate and the reactions they catalyze are mechanistically closely related. No such sequence identity was detectable, which suggested that ORF3 does not encode isochorismate synthase.

Analysis of the *entB*, *entA*, and *P15* gene products. The 3.25-kb *Hin*dIII-*Hin*dIII fragment (modified *Sac*II-*Xho*II fragment) was cloned into pT7-7 in both orientations (pJLT7-1A and pJLT7-1B). Plasmids pJLT7-2A and pJLT7-2B were constructed by subcloning the *Eco*RV-*Acc*I fragment containing ORF3 (*entA*) into the *Eco*RI site of pT7-7 in both orientations after modification of the *Acc*I end and addition of *Eco*RI linkers. Pulse-labeling with L-[^{35}S]methionine (28) followed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis allowed the proteins specified by each DNA strand to be identified (Fig. 5). pJLT7-1A was found to encode three polypeptides of apparent molecular weights 34,000, 25,000, and 15,000 (lane 1), whereas the complementary strand, cloned in pJLT7-1B, encoded no detectable protein sequences (lane 2). pJLT7-2A encoded a protein of molecular weight 25,000 (lane 3), but no product was detected from the opposite strand (pJLT7-2B, lane 4). These results are in agreement with the sizes predicted from the DNA sequence information.

Overproduction of the *entA* gene product. From recent genetic experiments (19), it has been suggested that the product of the *entA* gene could be a bifunctional enzyme with both diDHB-DH and isochorismate synthase activities. Although the sequence identity between ORF3 and ribitol dehydrogenase was suggestive of its proposed dehydrogenase function, the lack of sequence identity to anthranilate synthase and *para*-aminobenzoate synthase cast doubt on its putative isochorismate synthase function. We were thus prompted to overproduce and purify the *entA* gene product and assess its enzymatic activity.

To increase the level of the *entA* product within the cell, the gene was placed under the control of the *tac* promoter in pKK223-3. A method was developed, based on that described by Dale et al. (6), by which the 3′-exonuclease function of T4 DNA polymerase was used to specifically delete a region of single-stranded DNA between the restriction enzyme cleavage site created by one primer to the binding site of a second primer, creating a new *Eco*RI site in the process (J. Liu and C. T. Walsh, manuscript in preparation). In this way, the sequences upstream of *entA* were deleted to 19 base pairs before the initiation codon, leaving the natural *entA* Shine-Dalgarno sequence intact. Double-stranded DNA was produced and cloned into the *Eco*RI site of pKK223-3; the sequence of the vector-insert boundary in the resulting recombinant (pJLK-2) is shown in Fig. 6.

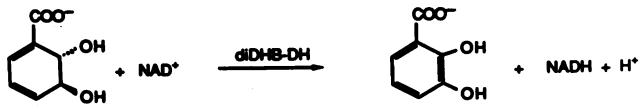

FIG. 7. Schematic of NAD-dependent conversion of diDHB to 2,3-dihydroxybenzoate.

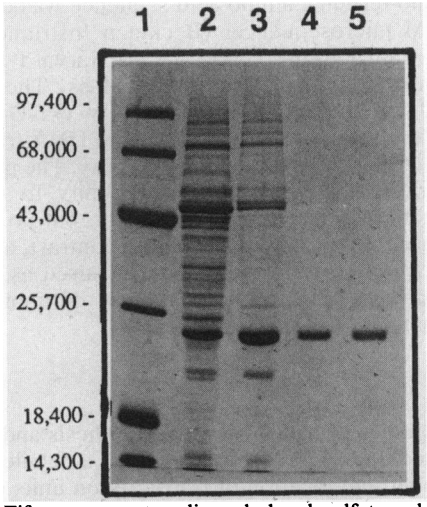

FIG. 8. Fifteen percent sodium dodecyl sulfate-polyacrylamide gel electrophoresis of samples taken at different stages of diDHB-DH purification. Lanes: 1, marker proteins; 2, crude cell extract; 3, 0 to 40% ammonium sulfate pellet after dialysis; 4, DEAE-Sepharose pool; 5, Blue A agarose pool. Molecular weight standards were as described in the legend to Fig. 5.

Purification of diDHB-DH. As noted in Materials and Methods, diDHB-DH can be assayed by the NAD-dependent conversion of the dihydroaromatic substrate diDHB to the aromatized catecholic product 2,3-dihydroxybenzoate (Fig. 7) by fluorometer analysis. A simple and rapid purification of diDHB-DH from the overproducing strain *E. coli* JM105(pJLK-2) was developed. The procedure consisted of precipitation with ammonium sulfate followed by chromatography on DEAE-Sepharose and Blue A agarose. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (Fig. 8) showed that the enzyme was essentially pure after ion-exchange column chromatography; the Blue A column removed a few contaminants. From 3 g of cell paste, approximately 2 mg of pure enzyme was obtained in 36% overall yield, which represented a 16-fold purification (Table 1). The final specific activity of the enzyme was 58 U mg^{-1}, allowing a turnover number of 1,520 min^{-1} to be calculated.

The native molecular weight of the enzyme was determined by chromatography on a Superose 12 gel filtration column. It was found to be 210,000, corresponding to the molecular weight of an octamer.

TABLE 1. Purification of diDHB-DH

Step	Vol (ml)	Activity (U ml^{-1})	Protein concn (mg ml^{-1})	Sp act (U mg^{-1})	Purification (fold)	Yield (%)
Crude extract	5	54	15	3.6	1	100
0–40% $(NH_4)_2SO_4$	5	39	1.6	24	7	72.2
DEAE	15	11.6	0.36	32	8.9	64.4
Blue A	15	6.4	0.11	59	16	35.6

The amino-terminal amino acid sequence was determined on an 890M microsequencer (Beckman Instruments, Inc., Fullerton, Calif.) by W. Lane at the Harvard University microchemistry facility, Cambridge, Mass. The sequence was found to be M-D-F-S-G-K-N-V-W-V-T-G-A, which confirmed the sequence predicted by the DNA translation.

Absence of isochorismate synthase activity. The preparation of pure diDHB-DH was tested for ability to synthesize isochorismate from substrate chorismate, but no isochorismate synthase activity was detected. In contrast, a crude cell extract of *Enterobacter aerogenes* 62-1, used as a positive control, showed detectable conversion of chorismate to isochorismate.

DISCUSSION

Recent studies of enterobactin biosynthesis and transport have focused on genetic aspects, such as a detailed analysis of the number and location of transcription units (8, 15, 23) and the regulation of gene expression in response to iron stress by the product of the *fur* regulatory protein (13, 20, 21). The enterobactin biosynthetic enzymes are interesting in that they catalyze a number of novel and stereochemically unusual reactions. Although these enzymes have been detected in crude extracts, no significant purification, characterization, or mechanistic analysis has been reported. This lack of progress has been due in part to the unavailability of the enzymes and their substrates and to an incomplete understanding of the components of the system. With this in mind, we initiated a study of reaction mechanisms by attempting to compare and contrast the reaction catalyzed by isochorismate synthase with those catalyzed by two other chorismate-utilizing enzymes, anthranilate synthase and *para*-aminobenzoate synthase, which had previously been studied in our laboratories (30). As a prelude to such a study, we attempted to clone and sequence the *E. coli* genes involved in synthesis of 2,3-dihydroxybenzoate in order to facilitate purification of the enzymes.

Genetic evidence for the locations of the *entA*, *entB*, and *entC* genes has in the past been somewhat controversial. Laird and Young (15) reported three separate transcription units within the *entCAGBE* cluster: *entCA*, *entB*, and *entE*, with *entG* expressed as either *entCAG* or *entGB*. Fleming et al. (8) reported a single operon comprising *entA(CGB)E* and were unable to resolve the order of the three central genes. Pickett et al. (23) later confirmed that *entA*, -B, -C, -E and -G were in a single transcription unit but did not comment further on gene order. Recently, Nahlik et al. (19) again suggested three transcription units, now established as *entE*, *entBG*, and *entAC*. DNA-sequencing studies reported here and by others (C. Earhart and M. A. McIntosh, personal communications) have finally clarified the situation.

The DNA sequence from the central portion of the cluster (Fig. 3) contains several ORFs. By comparison with complementation and in vitro and in vivo transcription-translation data, these have been assigned in order of transcription to *entE*, *entB*, *entA*, and a gene of unknown function, *P15*. There are little or no intergenic sequences, which leads to the conclusion that these genes are probably coexpressed on a single mRNA, from a promoter upstream of *entE* that is not present in our sequence. Indeed, the likely Shine-Dalgarno sequence necessary for translation of ORF3 is located within the coding region of ORF2; similarly, the ORF4 Shine-Dalgarno sequence is within ORF3. No sequences homologous to typical *E. coli* promoter −10 and −35 boxes were found in the cluster, which again suggests coexpression. If separate promoters are present, then they must also be located in coding regions, placing severe restraint on such sequences. A search was also made for sequences similar to the consensus sequence for the binding site of the *fur* product, which is normally located near the promoter of iron-regulated genes (4); no such sequences were identified.

The position or even existence of *entG* as a separate entity is now in doubt. *entG* has so far been defined only genetically; Nahlik et al. (19) were unable to identify the *entG* product in *E. coli* minicells, and deletion experiments defined a very small region of the cluster in which *entG* might lie. No gene was found in the position predicted for *entG*, between *entB* and *entA*. It is possible that the *entG* and *entB* mutations are allelic, as *entB* maps to the region now known to encode ORF2. The function of *P15* is also unknown. Because of the close linkage and likely coexpression with *entA*, *P15* is presumably also iron regulated and thus may also have a function in either the biosynthesis or uptake of enterobactin. As no mutants have been isolated in this gene, it is reasonable to propose that such mutations would have either no phenotype or a lethal phenotype; the latter is unlikely, since expression of this gene is not normally required for cell growth in an iron-rich environment.

On the basis of complementation of *E. coli entA* and *entC* mutants with a series of deletion derivatives of the *ent* gene cluster, Nahlik et al. (19) provided convincing evidence that both mutations were located within a 0.85-kb *Eco*RV-*Acc*I fragment and that both enzyme activities would be carried on a single 26,000-dalton polypeptide. It seemed somewhat surprising that two completely different enzyme activities would reside on a single polypeptide chain of this size. Also, Young et al. (33) demonstrated in crude extracts and initial purification work that isochorismate synthase activity could be separated from diDHB-DH and diDHB synthetase activity. ORF3 corresponds to the 26,000-dalton polypeptide, and the finding of the βαβ NAD-binding fold fingerprint was indicative of dehydrogenase function. However, the lack of sequence identity to *trpE* and *pabA* suggested that such a function was unlikely to be represented by isochorismate synthase. An assay for diDHB-DH that is based on detection of the reaction product 2,3-dihydroxybenzoate was developed. With diDHB as a substrate, activity was detected in cell extracts of a clone which overexpresses ORF3 under the control of the *tac* promoter (pJLK-2). The overproducing strain *E. coli* JM105(pJLK-2) produced diDHB-DH as 6% of total cell protein. DiDHB-DH was purified from this strain in a three-step procedure; amino-terminal sequence analysis confirmed the sequence of ORF3 predicted from the DNA sequence. This is the first reported purification to homogeneity of an enterobactin biosynthesis enzyme. The specific activity of the final active pool was 58 U mg^{-1}, the turnover number was 1,520 min^{-1}. By comparison, *E. coli* shikimate dehydrogenase, an enzyme involved in the biosynthesis of the chorismate precursor for the enterobactin pathway, has a turnover number of approximately 25,000 min^{-1} (5).

Gel filtration chromatography revealed that native diDHB-DH migrates at a position corresponding to an octamer. This is highly unusual; most *E. coli* dehydrogenases that have been studied, as well as two mechanistically related dehydrogenases from *Pseudomonas putida*, *cis*-benzene glycol dehydrogenase (1) and *cis*-toluene dihydrodiol dehydrogenase (24), are tetramers.

The purified enzyme preparation has no associated isochorismate synthase activity, and therefore the overall interpretation of the original complementation results was incorrect. This result leaves unanswered the question of the true

location of *entC*, but Earhart and McIntosh (personal communications) have now independently located *entC*; it was originally named *fepF* (22). DNA sequence analysis showed the anticipated sequence similarity to *trpE* and *pabB* and showed that *entC* is closely linked in the same transcription unit and promoter proximal to *entE*. An explanation of the complementation data may be that the mutation assigned as *entC* in the literature (36) actually lies within the *entA* structural gene (McIntosh, personal communication), which would imply that mutations in *entA* can in some subtle way affect isochorismate synthase activity, possibly by preventing formation of an active heterodimer (diDHB-DH or isochorismate synthase) or heteromultimeric complex. It will be interesting to purify isochorismate synthase and examine its activity and quaternary structure in the absence and presence of purified diDHB-DH.

ACKNOWLEDGMENTS

We are grateful to the following people for gifts and advice: Glenn Berchtold and Nina Quinn for diDHB, Ian Young for plasmids and *E. coli ent* strains, and Stan Tabor for modified T7 DNA polymerase and the T7 expression system. We are indebted to Charles Earhart and Mark McIntosh for communication of unpublished results and to Jack Chen for critical reading of the manuscript.

This work was supported by a Du Pont Graduate fellowship from Massachusetts Institute of Technology (to J.L.), a long-term fellowship from the European Molecular Biology Organization (to K.D.), and Public Health Service grant GM 20011 from the National Institutes of Health (to C.T.W.).

LITERATURE CITED

1. **Axcell, B. C., and P. J. Geary.** 1973. The metabolism of benzene by bacteria. Purification and some properties of the enzyme cis-1,2-dihydroxycyclohexa-3,5-diene (nicotinamide adenine dinucleotide) oxidoreductase (cis-benzene glycol dehydrogenase). Biochem. J. **136**:927–934.
2. **Bagg, A., and J. B. Neilands.** 1987. Molecular mechanism of regulation of siderophore-mediated iron assimilation. Microbiol. Rev. **51**:509–517.
3. **Bradford, M. M.** 1976. A rapid and sensitive method for the quantitation of protein utilizing the principle of protein-dye binding. Anal. Biochem. **72**:248–254.
4. **Calderwood, S. B., and J. J. Mekalanos.** 1987. Iron regulation of Shiga-like toxin expression in *Escherichia coli* is mediated by the *fur* locus. J. Bacteriol. **169**:4759–4764.
5. **Chaudhuri, S., I. A. Anton, and J. R. Coggins.** 1986. Shikimate dehydrogenase from *Escherichia coli*. Methods Enzymol. **142**: 315–320.
6. **Dale, R. M. K., B. A. McClure, and J. P. Houchins.** 1985. A rapid single-stranded cloning strategy for producing a sequential series of overlapping clones for use in DNA sequencing: application to sequencing the corn mitochondrial 18 S rDNA. Plasmid **13**:31–40.
7. **Earhart, C. F.** 1987. Ferrienterobactin transport in *Escherichia coli*, p. 67–84. *In* G. Winkelmann, D. van der Helm, and J. B. Neilands (ed.), Iron transport in microbes, plants and animals. VCH Verlagsgesellschaft, Weinheim, Federal Republic of Germany.
8. **Fleming, T. P., M. S. Nahlik, and M. A. McIntosh.** 1983. Regulation of enterobactin iron transport in *Escherichia coli*: characterization of *ent*::Mu d(Ap^r *lac*) operon fusions. J. Bacteriol. **156**:1171–1177.
9. **Fleming, T. P., M. S. Nahlik, J. B. Neilands, and M. A. McIntosh.** 1985. Physical and genetic characterization of cloned enterobactin genomic sequences from *Escherichia coli* K-12. Gene **34**:47–54.
10. **Gibson, F.** 1970. Chorismic acid. Biochem. Prep. **12**:94–97.
11. **Greenwood, K. T., and R. K. J. Luke.** 1976. Studies on the enzymatic synthesis of enterochelin in *Escherichia coli* K-12: four polypeptides involved in the conversion of 2,3-dihydroxybenzoate to enterochelin. Biochim. Biophys. Acta **454**:285–297.
12. **Greenwood, K. T., and R. K. J. Luke.** 1980. Studies on the enzymatic synthesis of enterochelin in *Escherichia coli* K-12, *Salmonella typhimurium*, and *Klebsiella pneumoniae*. Biochim. Biophys. Acta **614**:185–195.
13. **Hantke, K.** 1984. Cloning of the repressor protein gene of iron-regulated systems in *Escherichia coli* K-12. Mol. Gen. Genet. **197**:337–341.
14. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227**:680–685.
15. **Laird, A. J., and I. G. Young.** 1980. Tn*5* mutagenesis of the enterochelin gene cluster of *Escherichia coli*. Gene **11**:359–366.
16. **Lipman, D. J., and W. R. Pearson.** 1985. Rapid and sensitive protein similarity search. Science **227**:1435–1441.
17. **Loviny, T., P. M. Norton, and B. S. Hartley.** 1985. Ribitol dehydrogenase of *Klebsiella aerogenes*. Biochem. J. **239**:579–585.
18. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
19. **Nahlik, M. S., T. P. Fleming, and M. A. McIntosh.** 1987. Cluster of genes controlling synthesis and activation of 2,3-dihydroxybenzoic acid in production of enterobactin in *Escherichia coli*. J. Bacteriol. **169**:4163–4170.
20. **Neilands, J. B.** 1982. Microbial envelope proteins related to iron. Annu. Rev. Microbiol. **36**:285–309.
21. **Neilands, J. B., and K. Nakamura.** 1985. Regulation of iron assimilation in microorganisms. Nutr. Rev. **43**:193–197.
22. **Ozenberger, B. A., M. S. Nahlik, and M. A. McIntosh.** 1987. Genetic organization of multiple *fep* genes encoding ferric enterobactin transport functions in *Escherichia coli*. J. Bacteriol. **169**:3638–3646.
23. **Pickett, C. L., L. Hayes, and C. F. Earhart.** 1984. Molecular cloning of the *Escherichia coli* K-12 *entACGBE* genes. FEMS Microbiol. Lett. **24**:77–80.
24. **Rogers, J. E., and D. T. Gibson.** 1971. Purification and properties of *cis*-toluene dihydrodiol dehydrogenase from *Pseudomonas putida*. J. Bacteriol. **130**:1117–1124.
25. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
26. **Staden, R.** 1984. Computer methods to locate signals in nucleic acid sequences. Nucleic Acids Res. **12**:505–519.
27. **Staden, R., and A. D. McLachlan.** 1982. Codon preference and its use in identifying protein coding regions in long DNA sequences. Nucleic Acids Res. **10**:141–158.
28. **Tabor, S., and C. C. Richardson.** 1985. A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes. Proc. Natl. Acad. Sci. USA **82**:1074–1078.
29. **Tabor, S., and C. C. Richardson.** 1987. DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. Proc. Natl. Acad. Sci. USA **84**:4767–4771.
30. **Walsh, C. T., M. D. Erion, A. E. Walts, J. J. Delany III, and G. A. Berchtold.** 1987. Chorismate aminations: partial purification of *Escherichia coli*. PABA synthase and mechanistic comparison with anthranilate synthase. Biochemistry **26**:4734–4745.
31. **Wierenga, R. K., P. Terpstra, and W. G. H. Hol.** 1986. Prediction of the occurrence of the ADP-binding βαβ-fold in proteins, using an amino acid sequence fingerprint. J. Mol. Biol. **187**: 101–107.
32. **Woodrow, G. C., I. G. Young, and F. Gibson.** 1979. Biosynthesis of enterochelin in *Escherichia coli* K-12: separation of the polypeptides coded for by the *EntD,E,F*, and *G* genes. Biochim. Biophys. Acta **528**:145–153.
33. **Young, I. G., T. J. Batterham, and F. Gibson.** 1969. The isolation, identification and properties of isochorismic acid. An intermediate in the biosynthesis of 2,3-dihydroxybenzoic acid. Biochim. Biophys. Acta **177**:389–400.

34. **Young, I. G., and F. Gibson.** 1969. Regulation of the enzymes involved in the biosynthesis of 2,3-dihydroxybenzoic acid in *Aerobacter aerogenes* and *Escherichia coli*. Biochim. Biophys. Acta **177**:401–411.
35. **Young, I. G., L. M. Jackman, and F. Gibson.** 1969. The isolation, identification and properties of 2,3-dihydro-2,3-dihydroxybenzoic acid. An intermediate in the biosynthesis of 2,3-dihydrobenzoic acid. Biochim. Biophys. Acta **177**:381–388.
36. **Young, I. G., L. Langman, R. K. J. Luke, and F. Gibson.** 1971. Biosynthesis of the iron-transport compound enterochelin: mutants of *Escherichia coli* unable to synthesize 2,3-dihydroxybenzoate. J. Bacteriol. **106**:51–57.