JOURNAL OF BACTERIOLOGY, Feb. 1989, p. 775–783
0021-9193/89/020775-09$02.00/0
Copyright © 1989, American Society for Microbiology
Vol. 171, No. 2

Nucleotide Sequence of *Escherichia coli* Isochorismate Synthetase Gene *entC* and Evolutionary Relationship of Isochorismate Synthetase and Other Chorismate-Utilizing Enzymes

BRADLEY A. OZENBERGER,† TIMOTHY J. BRICKMAN, AND MARK A. McINTOSH*

Department of Microbiology, School of Medicine, University of Missouri-Columbia, Columbia, Missouri 65212

Received 5 August 1988/Accepted 2 November 1988

Biochemical analysis of the enzymatic activity catalyzing the conversion of chorismate to isochorismate in the enterobactin biosynthetic pathway attributed the reaction to the isochorismate synthetase enzyme, designated EntC. However, the lack of mutations defining this activity has hampered the precise identification of the *entC* structural gene. In this study, we engineered a stable insertion mutation into the chromosomal region between the enterobactin genes *fepB* and *entE*. This mutation disrupted the structural gene for a previously identified 44-kilodalton protein and eliminated production of 2,3-dihydroxybenzoic acid, the catechol precursor of enterobactin. The complete nucleotide sequence of this gene was determined and compared with the sequences of other genes encoding chorismate-utilizing proteins. The similarities observed in these comparisons not only indicated that the locus is *entC* but also supported the premise that these enzymes constitute a family of related proteins sharing a common evolutionary origin. In addition, in this and the accompanying paper (M. S. Nahlik, T. J. Brickman, B. A. Ozenberger, and M. A. McIntosh, J. Bacteriol. 171:784–790, 1989), evidence is presented indicating that the *entA* product is potentially a secondary factor in the chorismate-to-isochorismate conversion and that the prototypic *entC* lesion (*entC401*) resides in the structural gene for the EntA protein. Finally, polarity effects from the insertion mutation in *entC* on downstream biosynthetic genes indicated that this locus is the promoter-proximal cistron in an *ent* operon comprising at least five genes. Appropriate regulatory signals upstream of *entC* suggest that this operon is regulated by iron through interaction with the Fur repressor protein.

Biosynthesis of the catechol siderophore enterobactin in *Escherichia coli* is a two-stage process involving the initial production of 2,3-dihydroxybenzoic acid (DHBA) from the aromatic amino acid precursor chorismate and the subsequent conversion of DHBA and L-serine to the active chelator. The three soluble enzymes required for the initial stage are EntC, (isochorismate synthetase), EntB (2,3-dihydro-2,3-dihydroxybenzoate synthetase), and EntA (2,3-dihydro-2,3-dihydroxybenzoate dehydrogenase). Intermediates in the biosynthetic reactions catalyzed by these enzymes have been identified (32). The genes encoding the initial-stage enzymes were predicted to be linked transcriptionally on the clockwise end of the complex enterobactin gene cluster at min 13 on the *E. coli* chromosomal map (7). In the second stage of enterobactin biosynthesis, the proposed enterobactin synthetase multienzyme complex, suggested to contain the *entD*, *entE*, *entF*, and *entG* gene products (11), catalyzes the synthesis of one enterobactin molecule from three molecules each of DHBA and L-serine. These stage two genes are scattered throughout the enterobactin gene cluster (8, 19, 24, 25).

The enterobactin biosynthetic intermediate chorismate also serves as the pivotal substrate for several enzymes associated with key metabolic pathways (Fig. 1), including the synthesis of aromatic amino acids and vitamins, the folate coenzymes, and the substituted benzoic acids (e.g., DHBA). Enzymes that initiate these various metabolic pathways from the branch compound chorismate may therefore play a key regulatory role in cellular aromatic metabolism. In addition, these proteins may have common (and perhaps evolutionarily conserved) structural or functional features, such as those related to substrate recognition and binding, although their enzymatic mechanisms or cofactor requirements may be quite distinct. Evolutionary conservation of primary amino acid structure has been suggested for two such aminobenzoate synthetases, TrpE and PabB (9). To assess the primary structure of EntC and its potential relatedness to the other chorismate-utilizing enzymes, we sought to determine the nucleotide sequence of the *entC* structural gene.

In many early molecular studies defining the *ent*-region genes (14, 19, 25), no distinct genetic separation between *entC* and *entA* was established. This difficulty is a direct consequence of the fact that the same single lesions affecting either EntA or EntC functions were used in all of those studies. We recently established by genetic complementation studies that those mutant alleles, *entA403* and *entC401*, mapped to a single genetic locus defined as *entAC* (19). In the accompanying report (18), we established the nucleotide sequence and deduced amino acid residues of the 26-kilodalton (kDa) protein encoded by this locus. The primary amino acid sequence is characterized by clear relationships to established dehydrogenases, which suggests that these regions may correspond to the EntA activity of this protein; therefore, the protein was redefined as EntA. However, genetic evidence that the *entC401* lesion maps to the *entA* locus suggests that the 26-kDa protein functions in the

* Corresponding author.
† Present address: Department of Biology, Yale University, New Haven, CT 06511-8112.

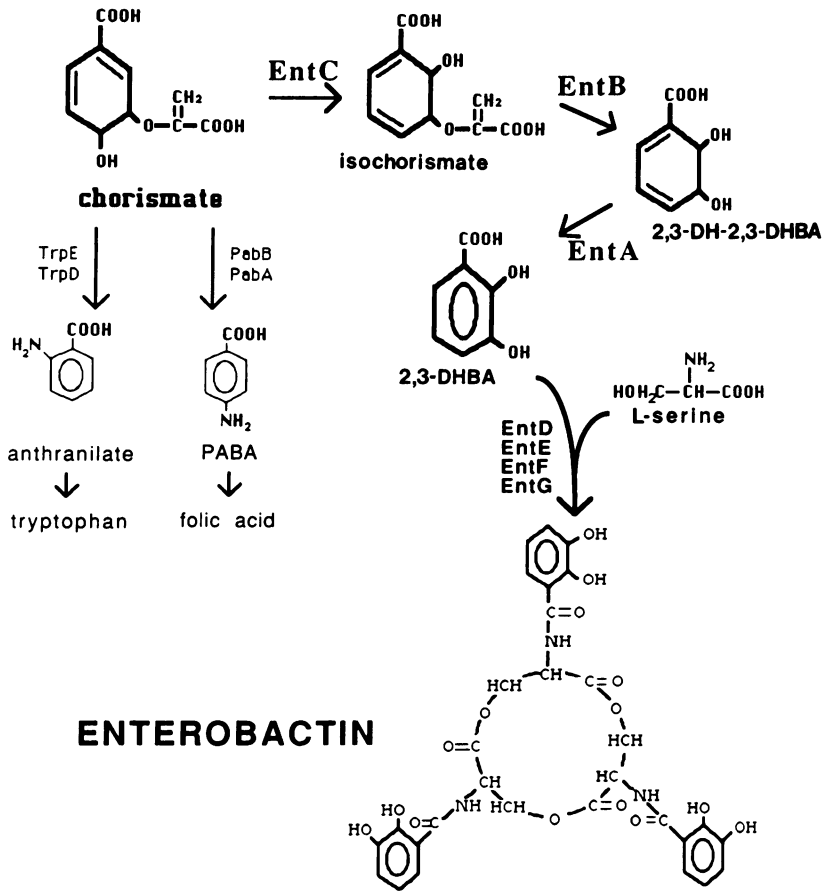

FIG. 1. Schematic showing that chorismate, a common precursor of enterobactin, anthranilate, and *p*-aminobenzoate (PABA), is converted to DHBA in a series of reactions involving three Ent enzymes; subsequently, DHBA and L-serine are used to synthesize the cyclic triester of dihydroxybenzoylserine (enterobactin). Four postulated Ent proteins make up the synthetase enzyme responsible for this final stage of synthesis (11). Conversion of chorismate to either anthranilate or PABA requires two gene products, TrpE and TrpD or PabB and PabA, respectively (9). Anthranilate is a precursor in the tryptophan biosynthetic pathway, and PABA is a component of folic acid. 2,3-DH-2,3-DHBA, 2,3-dihydroxy-2,3-dihydydroxybenzoate.

EntC-catalyzed reaction and cannot be presently overlooked.

In this study, we constructed a stable insertion mutation in the region between *fepB* and *entE* which we previously showed encoded a 44-kDa protein thought to be involved in enterobactin transport (22). This mutant strain was blocked in DHBA production and could not be restored phenotypically to wild type by transformation with cloned *entB* and *entA* genes. The nucleotide and predicted amino acid sequences of this locus show similarities to those of *trpE* and *pabB*, which suggests that the gene is *entC* and that it is part of a family of related genes encoding chorismate-utilizing enzymes. The *entC401* allele of the newly redefined *entA* locus may identify a region of interaction between EntA and EntC.

MATERIALS AND METHODS

Bacterial strains and plasmids. All *E. coli* strains used are described in Table 1. Recombinant plasmid constructions in

TABLE 1. Bacterial strains

Strain	Genotype	Reference or source
AB1515	*thi trpE purE proC leuB lacY mtl xyl rpsL azi fhuA tsx supA*	22
AN90	Same as AB1515 but *purE*$^+$ *entD*	22
AN191-60	Same as AB1515 but *purE*$^+$ *entC* Δ*recA*	19
AN193-60	Same as AB1515 but *purE*$^+$ *entA* Δ*recA*	19
HB101	*thi pro leu rpsL lac his gal recA*	*Escherichia coli* Genetic Stock Center, Yale University
MC4160	*araD* Δ(*lacIOPZYA*)U169 *rpsL thiA recA*	24
MM383	*polA12 thy rha lac rpsL*	17
MT147	*entC*::*kan* derivative of AB1515	This study
χ984	*minA minB pdxC purE his rpsL xyl ilv met*	22

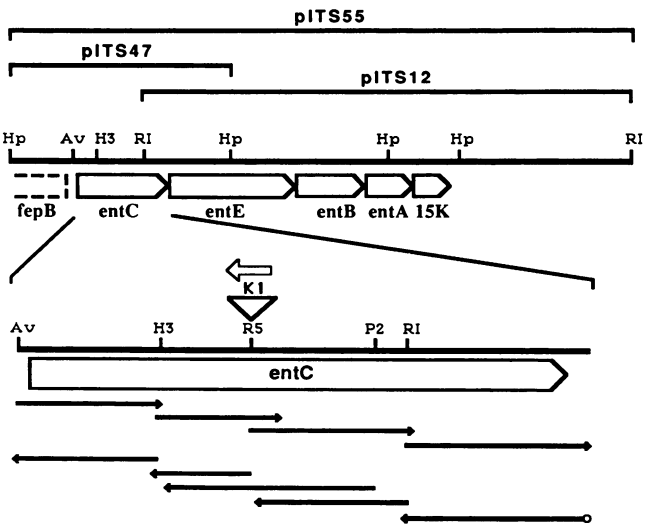

FIG. 2. Physical map of the right-hand region of the enterobactin system. The extents of the chromosomal DNA fragments inserted into each recombinant plasmid are indicated above the physical map. Constructions of pITS12 (19) and pITS47 (22) have been described, and that of pITS55 is detailed in the text. The *entC* gene encodes the 44-kDa protein (see text for explanation) and is enlarged to show the locations of the K1 insertion mutation and sequencing strategy. Insertion K1 consists of a 2.4-kilobase *Xho*I fragment from transposable element Tn*5* ligated into the *Eco*RV site in pITS47. The direction of transcription of the kanamycin resistance gene contained on the Tn*5* fragment (⇦) was determined by restriction enzyme cleavage analysis. ⟶, Region of nucleotide sequence determined from the specified restriction site. The circle on the bottom arrow represents the priming site for a synthetic oligonucleotide (designated CR1.3) used to obtain the nucleotide sequence of the template strand near the 3′ end of *entC*. Abbreviations for restriction enzyme sites: Av, *Ava*I; H3, *Hin*dIII; Hp, *Hpa*I; P2, *Pvu*II; RI, *Eco*RI; R5, *Eco*RV.

the vector pBR328 (31) are designated pITS (*p*lasmids containing *i*ron *t*ransport *s*ystem genes) and are described in Fig. 2. Fragments isolated for DNA sequence analysis were cloned into the polylinker region of pUC18 and pUC19 (21) or pGEM3Z (Promega Biotec, Madison, Wis.).

Media, chemicals, and enzymes. Tris-glucose minimal medium containing 200 μM 2,2′-dipyridyl was used as the selective iron-depleted medium. The concentrations of supplements added have been previously described (7, 15). Chrome azurol S (CAS) agar for the detection of siderophores was made by the procedure of Schwyn and Neilands (29) and supplemented with 100 μM 2,2′-dipyridyl. Luria-Bertani (LB) broth was used as complete medium (16). Ampicillin (Ap) and kanamycin (Km) were used at a concentration of 50 μg/ml. Restriction and modification enzymes were purchased from Promega Biotec; Sequenase was purchased from United States Biochemical Corp., Cleveland, Ohio.

Plasmid isolation and transformation. Plasmids were isolated by the alkaline lysis procedure of Birnboim and Doly (4). $CaCl_2$-treated cells were transformed as described by Cohen et al. (5).

Minicell isolation and protein analysis. Strain χ984 was transformed with the various recombinant plasmids, and minicells were isolated by sucrose gradient centrifugation (22). Minicell preparations were labeled with [^{35}S]methionine, and protein samples were solubilized and examined by electrophoresis on sodium dodecyl sulfate-polyacrylamide gels as described previously (22).

Marker exchange. The Km^r marker from pITS47-K1 (Fig. 2) was transferred to the chromosome of MM383 by homologous recombination mediated by the flanking *entC* DNA sequences. MM383 [*polA*(Ts)] transformants containing pITS47-K1 were passaged three times at 42°C, and Km^r Ap^s survivors were selected as described previously (22). These recombinants were examined for enterobactin production on CAS agar (29) or by bioassay (7) and for transport by cross-feeding with purified enterobactin. The *entC*::*kan* allele from one of these strains was transduced into strain AB1515 with phage P1 (16) to produce strain MT147.

DNA sequence determination and analysis. DNA sequences were obtained by the dideoxy-chain termination method of Sanger et al. (28). DNA fragments were cloned into pUC18, pUC19, or pGEM3Z to permit sequence determination on both DNA strands. The sequencing strategy is depicted in Fig. 2. Plasmids to be used as sequencing templates were isolated by the alkaline lysis procedure (4) and precipitated with 6.5% (final concentration) polyethylene glycol in the presence of 0.8 M NaCl. Double-stranded template was denatured in 0.2 M NaOH–0.2 mM EDTA (pH 7.0), precipitated, and suspended in sequencing buffer (40 mM Tris [pH 7.5], 20 mM $MgCl_2$, 50 mM NaCl). Primer was annealed at 37°C for 15 min. Buffer, nucleotide mixtures, formamide stop solution, and modified T7 DNA polymerase were supplied in Sequenase DNA sequencing kits purchased from United States Biochemical Corp. Oligonucleotide primers were purchased from the University of Missouri-Columbia DNA Core Facility and synthesized on a model 380A DNA synthesizer (Applied Biosystems, Inc., Foster City, Calif.). Sequencing reactions were electrophoresed on 0.4- to 0.8-mm 8% polyacrylamide wedge gels, which were then exposed to X-Omat XRP5 film (Eastman Kodak Co., Rochester, N.Y.) at −70°C for variable times. Sequence data were analyzed by using the Microgenie software programs (26) distributed by Beckman Instruments, Inc., Palo Alto, Calif.

RESULTS

Construction of an insertion mutation between *fepB* and *entE*. We previously showed in genetic complementation studies that the *entC401* and *entA403* mutations present in strains AN191 and AN193, respectively, represent independent alleles of the same cistron (19), providing a rationale for the previous difficulty in genetic separation of these mutations (14). This genetic locus was designated *entAC* (19) because of its apparent involvement in both enzymatic activities and was shown to encode a 26-kDa protein product. However, the deduced amino acid sequence of this protein (18) did not display strong similarities to those of other enzymes that utilize chorismate as a substrate. As a consequence, we concluded that it was not the major enzymatic component of isochorismate synthetase and redefined its gene as *entA* (18). We therefore focused our attention on other proteins encoded in this genetic region for evidence of involvement in DHBA production and similarities to other chorismate-modifying enzymes.

The phenotypically uncharacterized 15-kDa polypeptide encoded by a genetic locus downstream of *entA* (designated *P15* for convenience) was confirmed by nucleotide sequence analysis (18), but it also contained no such homologies. In addition, an insertion mutation in its structural gene had no obvious phenotypic effects on enterobactin synthesis or transport (T. J. Brickman, unpublished results).

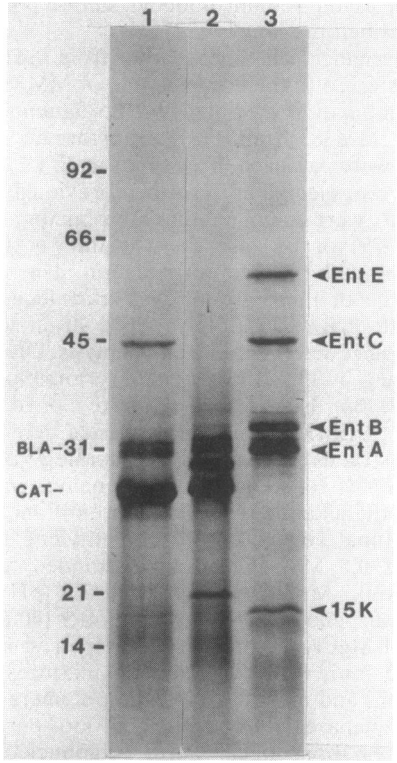

FIG. 3. Autoradiogram of ^{35}S-labeled proteins expressed from minicells. Minicells containing the specified plasmids were isolated and labeled as described in Materials and Methods. Proteins were separated on a 7.5 to 18% continuous gradient polyacrylamide–3% urea gel in the presence of sodium dodecyl sulfate. The positions of *ent* products are indicated on the right. Positions of β-Lactamase (BLA) and chloramphenicol acetyltransferase (CAT) expressed by the vector are also indicated. Migration positions of molecular size standards expressed in kilodaltons are shown on the left. Lanes: 1, pITS47; 2, pITS47-K1; 3, pITS55.

The remaining protein of undocumented function in this genetic region was the 44-kDa polypeptide expressed from the chromosomal sequences between *fepB* and *entE* (22). On the basis of initial investigation of a Tn*5* insertion mutation in the vicinity of its structural gene, we hypothesized that this protein was involved in enterobactin transport and preliminarily designated its structural gene *fepF*. However, detailed analysis of the insertion by Southern hybridization revealed that it was not stably integrated in this chromosomal region, a finding that failed to support that premise.

We constructed a stable insertion in the gene specifying the 44-kDa protein by cloning a Km^r cassette into the single *Eco*RV site of the recombinant plasmid pITS47, which contains the chromosomal DNA fragment extending from within the *fepB* gene to the *entE* gene (Fig. 2) and expresses the 44-kDa protein (Fig. 3) (22). A 2.4-kilobase *Xho*I restriction fragment which confers resistance to kanamycin was isolated from a plasmid carrying the transposable element Tn*5* (3). The *Xho*I ends were filled in with appropriate deoxynucleotide triphosphates, using the large fragment of DNA polymerase (Klenow), and the repaired DNA fragment was then ligated into pITS47 at the *Eco*RV site to create pITS47-K1 (Fig. 2). This insertion completely disrupted synthesis of the 44-kDa protein (Fig. 3, lane 2). pITS47-K1 was transformed into strain MM383 [*polA*(Ts)], and after several passages at the nonpermissive temperature, recombinants that had lost the plasmid Ap^r marker but maintained the Km^r marker were selected. The Km^r marker was then transduced to strain AB1515, and the mutant strain was designated MT147 (Table 1). The chromosomal location of the Km^r marker was confirmed by Southern hybridization.

TABLE 2. Genetic complementation[a]

Plasmid	Complementation with:		
	MT147 (*entC147*)	AN191-60 (*entC401*)	AN193-60 (*entA403*)
pITS47	±	−	−
pITS47-K1	−	−	−
pITS12	−	+	+
pITS55[b]	++	++	++
pITS56	−	+	+

[a] Each recipient strain was transformed with the indicated plasmid and examined for complementation, measured as the ability (+) or inability (−) to produce siderophore as detected on CAS agar (29). The ± phenotype indicates an intermediate level of siderophore release as compared with the positive result of AN193-60(pITS12).

[b] Complementation with pITS55 is designated ++ to indicate much greater siderophore production than in the control strain, AN193-60(pITS12). This increased production can be explained by the presence of the iron-responsive control element upstream of *entC* (defined in the accompanying report [18]) on the cloned DNA fragment represented by pITS55 (see Fig. 2). pITS55 also complements *entE* and *entB* mutants.

Characterization of MT147. MT147 was examined for enterobactin biosynthesis and transport phenotypes. The mutant was unable to grow on iron-depleted media unless the media were supplemented with purified enterobactin. Closer examination for a transport phenotype in a ^{55}Fe-enterobactin uptake assay revealed little or no effects on ferric enterobactin transport (data not shown). However, examination of MT147 in a cross-feeding bioassay (7) and analysis on CAS agar (29) revealed that it produced no detectable enterobactin. Siderophore biosynthesis was restored to a low level when MT147 was grown on CAS agar supplemented with purified DHBA, which indicated that the mutant was blocked in the synthesis of DHBA from chorismate. These data can be explained by the involvement of the 44-kDa protein in the initial stage of enterobactin production or by lack of expression of downstream, transcriptionally linked *ent* genes caused by the *kan* insertion. Since the inserted *kan* gene was shown to be transcribed in the direction opposite that of the disrupted gene, polarity effects on any downstream gene might be strong.

The Ent$^-$ phenotype of MT147 was further analyzed by genetic complementation studies. MT147 was transformed with various recombinant plasmids and examined for enterobactin biosynthesis on CAS agar relative to levels produced by similar transformants of the mutant strains AN191-60 (*entC401*) and AN193-60 (*entA403*) (Table 2). Each of the point mutants was complemented by pITS12 (Fig. 2) but not pITS47; pITS12, which does not express the 44-kDa protein (19), did not complement MT147. pITS47, which encodes a wild-type 44-kDa protein (Fig. 3), complemented the lesion in MT147 only weakly. pITS47-K1 was unable to restore enterobactin synthesis in MT147 to any degree. These data suggest that the insertion in MT147 produces strong polarity effects on enterobactin biosynthesis genes downstream of its location, thereby precluding wild-type expression of those genes and decreasing the apparent genetic complementation of MT147 by pITS47. In addition, the failure of pITS12 (which expresses *entB* and *entA*) to restore DHBA production at any level when introduced into MT147 suggests that the insertion defines a gene involved in DHBA synthesis,

presumably encoding a component of isochorismate synthetase (EntC) activity.

To overcome the polarity effects of the mutation in MT147, a plasmid was constructed which carries all of the known *ent* loci in this genetic region. The DNA fragment in pITS47 which extends from the *Eco*RI site in the gene encoding the 44-kDa protein (EntC) rightward to the *Eco*RI site in the vector pBR328 was replaced with the 6.6-kilobase *Eco*RI fragment from pITS12 to engineer pITS55 (Fig. 2). This plasmid contains the chromosomal DNA from the *fepB* gene rightward to a region beyond the known enterobactin genes. In conjunction with the construction of pITS55, a clone was isolated with the 6.6-kilobase *Eco*RI fragment in the wrong orientation. This clone was designated pITS56.

pITS55 complemented MT147, AN191-60, and AN193-60 (Table 2). In fact, the halos resulting from enterobactin release on CAS agar were much larger with MT147(pITS55) and AN191-60(pITS55) or AN193-60(pITS55) than with AN191-60(pITS12) or AN193-60(pITS12). pITS56 complemented AN191-60 and AN193-60 to the same degree as did pITS12 and was unable to complement MT147. These data suggest that the contiguous genes beginning with *entC* and extending rightward to the *P15* gene are cotranscribed from a control region upstream of the former gene (Fig. 2). This supposition is confirmed by the fusion studies described in the accompanying paper (18). These conclusions suggest that expression of the *ent* genes contained on pITS12 is not under normal control signals but may be directed by uncharacterized vector or low-level secondary promoters.

Proteins expressed from pITS55 in *E. coli* minicells were examined by sodium dodecyl sulfate-polyacrylamide gel electrophoresis. EntE, EntB, EntA (molecular weights of 58,000, 32,500, and 26,000, respectively [19]) and both the 15- and 44-kDa proteins were detected (Fig. 3, lane 3).

Nucleotide sequence of *entC*. Subclones of pITS47 were constructed in pUC vectors to incorporate techniques for sequencing from double-stranded templates. DNA fragments from pITS47, generated by restriction nucleases *Ava*I, *Hin*dIII, *Eco*RV, *Pvu*II, and *Eco*RI, were used for sequence determination in both directions from appropriate restriction sites (Fig. 2). A synthetic oligonucleotide (designated CR1.3) was used to obtain the sequence from the 5′ end of *entE* toward the *Eco*RI site in *entC* (Fig. 2). The nucleotide sequence from the *Ava*I site just upstream of this gene to the end of the proposed coding region is presented in Fig. 4. The sequence of this 1,400-base-pair region was determined unambiguously on both strands of the DNA.

Examination of all translational reading frames identified one which was open for 1,173 bp. This reading frame initiated with an ATG codon at base 121 preceded 5 bases by a polypurine sequence that bore strong resemblance to a consensus Shine-Dalgarno ribosome-binding site (30); the sequence encoded a protein of 391 amino acids. The calculated molecular weight of this polypeptide was 42,917, in agreement with the observed migration of the major product of pITS47 by gel electrophoresis (Fig. 3, lane 1).

Comparison of *entC* with *trpE* and *pabB*. Genetic analysis of mutant MT147 indicated that the 44-kDa protein acts in the biochemical conversion of chorismate to DHBA and is likely the major protein component of the EntC enzyme. Further support for this suggestion was obtained by direct comparison of its primary amino acid sequence with those of two other enzymes which utilize chorismate as a substrate and which are encoded by the *trpE* and *pabB* genes. *trpE* encodes anthranilate synthetase component I (20), and *pabB* encodes *p*-aminobenzoate synthetase component I (9) (Fig. 1). Alignment of the *entC* product with TrpE and PabB revealed extensive similarities concentrated at the carboxy end of each protein (Fig. 5). It is emphasized that these amino acid alignments are optimized with the introduction of a minimum number of gaps to maintain those regions previously identified (9) as evidence for evolutionary relatedness among this family of proteins. The addition of EntC to this analysis corroborates and extends the previous comparisons. The amino termini of these proteins did not display such widespread conservation of sequence. At the nucleotide level in the compared regions, *entC* showed 46% homology to *trpE* and 45% homology to *pabB*, whereas *trpE* and *pabB* are 40% similar (9). The amino acid sequences in this same region also showed greater than 40% similarity. No significant amino acid similarities were found between EntC and two other aromatic pathway enzymes, PheA and TyrA (12).

Comparison of codon usage in these three proteins (data not shown) indicated a high level of similarity; in addition, the patterns of codon usage by these proteins were consistent with those of *E. coli* genes that are not generally highly expressed (10). The preferential use of codons in the *entC* gene that utilize the more prevalent tRNA molecules in *E. coli* cells (13) further supports the proposed translational reading frame.

Putative regulatory sequences. The putative −35 and −10 sequences controlling transcription initiation are underlined in Fig. 4, and the proposed transcription initiation site (+1) is denoted (T. J. Brickman and B. A. Ozenberger, unpublished data). The −35 sequence (TTGACA) is perfectly homologous to the consensus sequence (27), and the −10 sequence (TAGGTT), although less conserved, contains the critical T, A, and T residues at positions 1, 2, and 6, respectively. A nucleotide sequence that strongly resembles the Fur repressor-binding site (6) is also present between +1 and the predicted translational start codon at +55. The Fur protein acts as a repressor for numerous iron-regulated genes (1) and has been shown to bind DNA in the presence of divalent cations such as ferrous iron (2). Expression of *entC* is shown in the accompanying paper (18) to be iron regulated.

DISCUSSION

We have established the complete nucleotide sequence of a region of more than 1,400 bp separating the established genetic loci *fepB* and *entE* in the enterobactin gene cluster. This sequence is characterized by a 1,173-bp open reading frame encoding a 391-residue polypeptide with a calculated molecular weight of 42,917, in agreement with the previous estimate of 44,000 (22) derived by gel electrophoresis. Phenotypic effects from an insertion mutation, constructed in vitro in the cloned gene sequence and recombined into the chromosomal locus, established that this protein is a component of the enzymatic machinery involved in the conversion of the aromatic compound chorismate to DHBA, the catechol moiety of enterobactin. The considerable similarities evident in the amino acid sequences of this protein and the chorismate-binding components (TrpE and PabB) of the enzymes anthranilate synthetase and *p*-aminobenzoate synthetase led to the conclusion that it is the major enzymatic component of isochorismate synthetase, EntC.

It has been proposed that the group of enzymes which direct the common substrate chorismate into a variety of metabolic pathways may be encoded by a family of evolutionarily related genes which have since diverged to produce

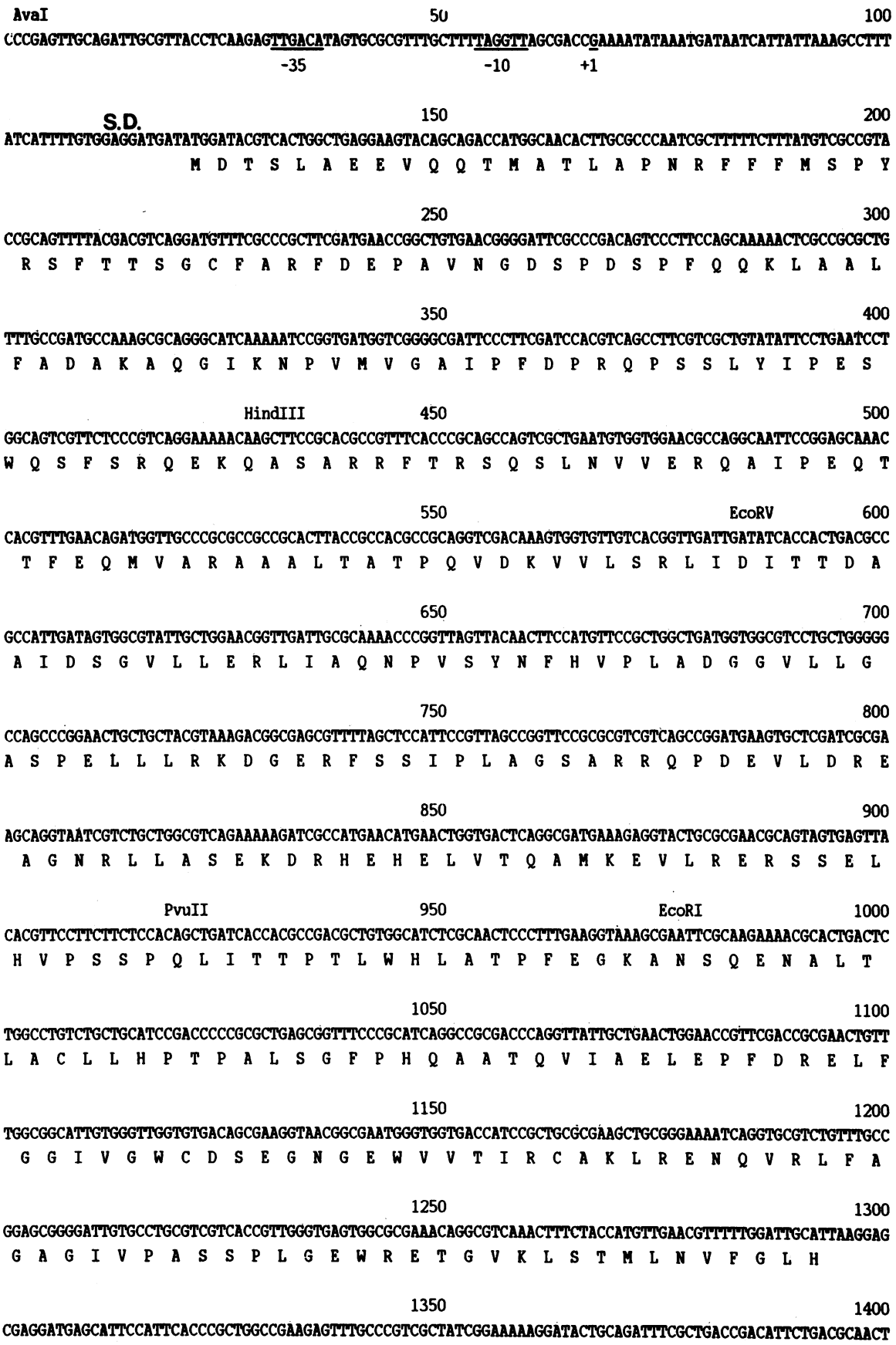

FIG. 4. Nucleotide and amino acid sequence of the *E. coli entC* locus. Nucleotides representing the coding strand are depicted 5′ to 3′ beginning at the *Ava*I site located between the *fepB* and *entC* genes and ending beyond the proposed *entC* translated sequences. Purported transcription initiation regulatory sequences are underlined. The predicted ribosome-binding site (S.D.) is identified five bases upstream of the translation start codon. Significant restriction enzyme recognition sites are identified. The K1 insertion mutation (see legend to Fig. 2) is in the center of the *Eco*RV site at base 585. The *Eco*RI site at base 975 represents the left end of the DNA fragment from pITS12.

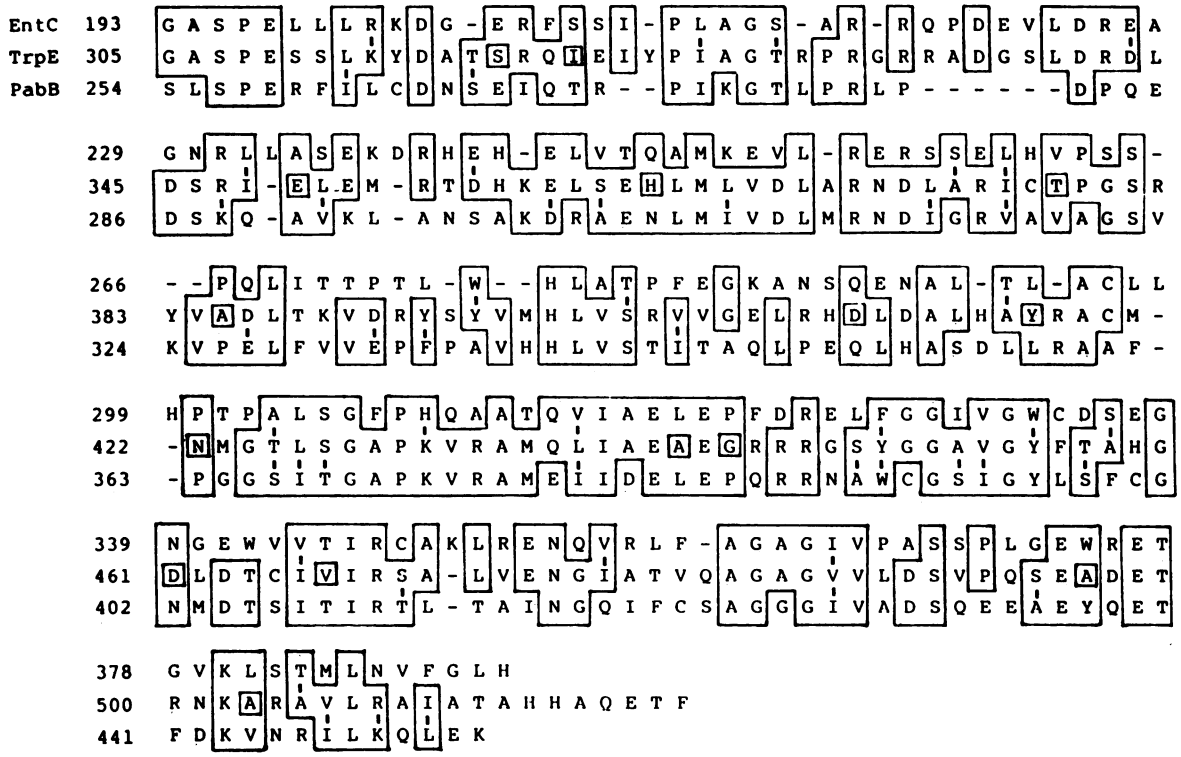

FIG. 5. Amino acid alignment of segments of EntC, TrpE, and PabB. The first amino acid in each line is numbered to indicate its position in the respective protein. Similar but nonidentical amino acids are denoted by a line between two sequences. Gaps in the sequences are indicated by hyphens.

the present variation in reaction end products (9). This premise is substantiated by the facts that two of these genes, *trpE* and *pabB*, show 40% similarity at the nucleotide level and their products are 26% similar. The concept is further reinforced by the current findings that *entC* is 45 to 46% homologous to these same genes and that at the amino acid level, across the entire coding length, the proteins are 19 to 26% similar. The greatest regions of similarity are found in the carboxy-terminal halves of these proteins, where the residue similarities approach or exceed 40%, which indicates that these regions may define components involved in chorismate interaction. These observations provide considerable support for the hypothesis (9) that these enzymes are encoded by a family of genes that have a common evolutionary origin. The amino ends of these proteins are characterized by relatively few similarities, which suggests considerable divergence that might be related to such parameters as cofactor specificity, enzymatic mechanism, and variability in response to important metabolic signals (e.g., feedback inhibition). In addition, it was proposed that the anthranilate synthetase enzyme lacked tryptophan residues and that other chorismate-utilizing enzymes would contain this amino acid so that aromatic metabolism would be diverted to tryptophan production under tryptophan starvation conditions (9, 20). The higher than average concentration of such residues in EntC further substantiates that concept.

Two of the aromatic pathway enzymes, anthranilate synthetase and *p*-aminobenzoate synthetase, have been shown to catalyze similar reactions involving chorismate and glutamine and produce similar products as a result of amidotransferase reactions that differ in the position to which the amide substituent is transferred to the benzene ring. The two enzymes are also structurally similar, each consisting of two nonidentical subunits; component I is responsible for chorismate binding, and component II exhibits the amidotransferase activity (9). The EntC component of isochorismate synthetase is expected to be related to component I of these enzymes, since chorismate is the common substrate for all three reactions. However, since isochorismate is produced by the transfer of a hydroxyl rather than an amide group to the benzene ring, no function analogous to the amidotransferase activity (component II) of these enzymes would be required. As expected, no similarities between EntA or EntB and TrpD or PabA (which correspond to component II of anthranilate synthetase and *p*-aminobenzoate synthetase, respectively) were identified (18).

It is important, however, that the original genetic lesion defining EntC activity, *entC401*, maps to the *entA* locus (19), which may suggest that EntC also is composed of two nonidentical subunits. In this regard, a small region of similarity to TrpE and PabB was identified by computer analysis of the EntA peptide sequence (data not shown); although the importance of this observation is not clear, it may reflect an additional component of substrate binding that is missing in the EntC protein (which at 391 residues is much smaller than the 453-residue PabB and 520-residue TrpE proteins). Alternatively, EntA may contain an enzymatically active component that lends substrate or mechanistic specificity to the EntC enzyme and clearly distinguishes it from the other pathways. Further mutational and biochemical analyses are required to clarify the potential EntA-EntC interaction.

This study established that the *entC*::*kan* mutant strain MT147 is only weakly restored to enterobactin production by the presence of a wild-type *entC* gene on a multicopy vector and that full synthesis capabilities in this strain

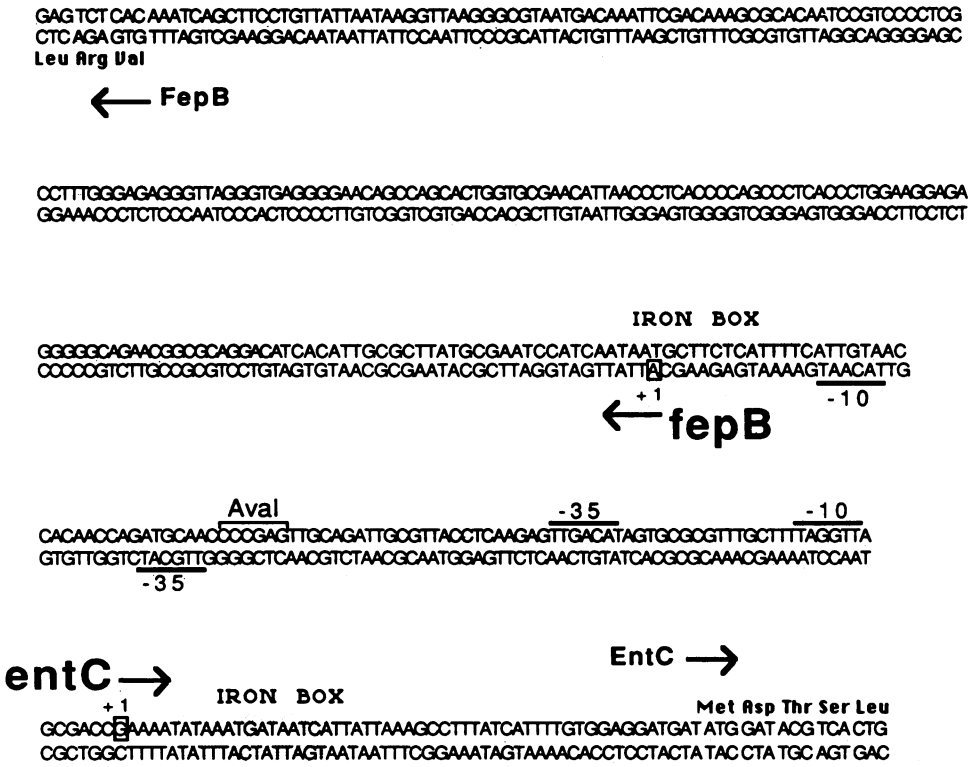

FIG. 6. Nucleotide sequence of the *fepB-entC* intercistronic region. The DNA sequence of the 394-bp region from within the translated region of *fepB* to that of *entC* is depicted. The designated *Ava*I recognition site corresponds to that identified in Fig. 4. Identified transcriptional initiation signals (−10 and −35) for both transcripts are based on their relationship to delineated +1 transcript start sites (Brickman and Ozenberger, unpublished data). The proposed translation start site for EntC is identified in Fig. 4, and that for FepB is from the work of M. Elkins and C. F. Earhart (Abstr. Annu. Meet. Am. Soc. Microbiol., K93, p. 218). The shaded areas correspond to potential Fur recognition sequences (1, 6).

require *entC* and all the *ent* genes downstream of this locus. These observations, combined with the DNA sequence and gene fusion data presented in the accompanying report (18), clearly establish that the right-hand *ent* gene cluster is organized as an operon containing at least five genes in the order *entC*, *entE*, *entB*, *entA*, and *P15* and that iron-mediated expression of this polycistronic transcript is controlled by the DNA sequences upstream of *entC*. Previous results (14, 19, 25) suggesting that these cistrons are independently transcribed can be explained by the fact that the low-level expression of independent *ent* genes on multicopy plasmids or downstream of transposon insertions that were not absolutely polar were measured qualitatively in the sensitive bioassays or by growth under iron-depleted conditions in those experiments. This interpretation is reinforced by quantitative comparisons of enterobactin production from *entB* and *entA* mutant strains harboring pITS55, which contains the entire operon, and pITS12, which does not include the major control sequences. Siderophore synthesis can clearly be measured in the pITS12 derivatives but is weak compared with production when the natural control signals are present.

From the DNA sequence data described in Fig. 4 and additional sequencing experiments not described, we have established the primary nucleotide sequence of the almost 400 bp separating the divergently transcribed *fepB* and *entC* genes (Fig. 6). By primer extension analysis, we have preliminarily identified the *entC* transcription start site (+1) some 55 bp upstream of the proposed EntC translation initiation codon (Brickman and Ozenberger, unpublished results). Within this region there is a sequence very similar to the consensus Fur-binding site (6) which, if confirmed by DNA-binding experiments, would contribute to the iron-regulated expression of the *entCEBA* (*P15*) operon. Upstream of the *entC* +1 site and on the opposite strand is the similarly identified initiation nucleotide for the *fepB* transcript. In the 103 bp separating these divergent mRNAs are located the sequences resembling canonical −10 and −35 promoter elements for both messages. Another sequence that may bind the Fur repressor is predicted to overlap the *fepB* transcript initiation site. Unlike the *entC* transcript, which has a small (55-nucleotide) leader RNA, the *fepB* mRNA appears to extend more than 200 nucleotides before the predicted FepB translation start codon. Structural analysis suggests that potential for extensive regions of secondary structure in this RNA molecule that may play a role in modulating *fepB* expression by mechanisms that act subsequent to initiation of transcription by RNA polymerase in the absence of bound Fur repressor. A detailed molecular analysis of this control region will be published elsewhere (T. J. Brickman, B. A. Ozenberger, and M. A. McIntosh, manuscript in preparation).

A similar but distinct transcriptional organization occurs on the left end of the enterobactin gene cluster between the *fepA* and *fes* genes (23, 24). This intercistronic region also produces diverging transcripts, but their promoters are extensively overlapping, with the primary initiation sites only 18 bp apart. Probable Fur-binding sequences are also apparent in this region. Both the *fepA* and *fes* transcripts are characterized by leader RNA lengths of more than 125

nucleotides, with the *fepA* mRNA, like the *fepB* mRNA, capable of potential base pairing. These observations taken together suggest that multiple control mechanisms are involved both before and after transcription initiation to modulate and coordinate the complex array of enterobactin biosynthesis and transport genes from these two bidirectional control sequences. Elucidation of these regulatory processes will require extensive molecular characterization of mutations directed to these central regions.

ACKNOWLEDGMENTS

We are indebted to Brian Nichols for suggestions and discussions concerning the potential structural relationships among chorismate-utilizing proteins and to M. Elkins for communication of unpublished sequence data. We thank G. S. Pettis, M. S. Nahlik, and C. F. Earhart for helpful discussions during the course of these studies. We thank Juanita Neuner for preparation of the manuscript.

This work was supported by grant DMB 8416017 from the National Science Foundation. T.J.B. is the recipient of Public Health Service predoctoral traineeship 1 T32 AI07276 from the National Institute of Allergy and Infectious Diseases.

LITERATURE CITED

1. **Bagg, A., and J. B. Neilands.** 1987. Molecular mechanism of regulation of siderophore-mediated iron assimilation. Microbiol. Rev. **51**:509–518.
2. **Bagg, A., and J. B. Neilands.** 1987. Ferric uptake regulation protein acts as a repressor, employing iron(II) as a cofactor to bind the operator of an iron transport operon in *Escherichia coli*. Biochemistry **26**:5471–5477.
3. **Berg, D. E., M. A. Schmandt, and J. B. Lowe.** 1983. Specificity of transposon Tn*5* insertion. Genetics **105**:813–828.
4. **Birnboim, H. C., and J. Doly.** 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. **7**:1513–1523.
5. **Cohen, S. N., A. C. Y. Chang, and L. Hsu.** 1972. Nonchromosomal antibiotic resistance in bacteria: genetic transformation of *Escherichia coli* R-factor DNA. Proc. Natl. Acad. Sci. USA **69**:2110–2114.
6. **De Lorenzo, V., S. Wee, M. Herrero, and J. B. Neilands.** 1987. Operator sequences of the aerobactin operon of plasmid ColV-K30 binding the ferric uptake regulation (*fur*) repressor. J. Bacteriol. **169**:2624–2630.
7. **Fleming, T. P., M. S. Nahlik, and M. A. McIntosh.** 1983. Regulation of enterobactin iron transport in *Escherichia coli*: characterization in *ent*::Mu d(Ap^r *lac*) operon fusions. J. Bacteriol. **156**:1171–1177.
8. **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.
9. **Goncharoff, P., and B. P. Nichols.** 1984. Nucleotide sequence of *Escherichia coli pabB* indicates a common evolutionary origin of *p*-aminobenzoate synthetase and anthranilate synthetase. J. Bacteriol. **159**:57–62.
10. **Grantham, R., C. Gautier, M. Gouy, M. Jacobzone, and R. Mercier.** 1981. Codon catalog usage is a genome strategy modulated for gene expressivity. Nucleic Acids Res. **9**:r43–r74.
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. **Hudson, G. S., and B. E. Davidson.** 1984. Nucleotide sequence and transcription of the phenylalanine and tyrosine operons of *Escherichia coli* K-12. J. Mol. Biol. **180**:1023–1051.
13. **Ikemura, T.** 1981. Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the *E. coli* translational system. J. Mol. Biol. **151**:389–409.
14. **Laird, A. J., and I. G. Young.** 1980. Tn*5* mutagenesis of the enterochelin gene cluster of *Escherichia coli*. Gene **11**:359–366.
15. **McIntosh, M. A., and C. F. Earhart.** 1977. Coordinate regulation by iron of the synthesis of phenolate compounds and three outer membrane proteins in *Escherichia coli*. J. Bacteriol. **131**:331–339.
16. **Miller, J. H.** 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
17. **Monk, M., and J. Kinross.** 1972. Conditional lethality of *recA* and *recB* derivatives of a strain of *Escherichia coli* K-12 with a temperature-sensitive DNA polymerase I. J. Bacteriol. **109**: 971–978.
18. **Nahlik, M. S., T. J. Brickman, B. A. Ozenberger, and M. A. McIntosh.** 1989. Nucleotide sequence and transcriptional organization of the *Escherichia coli* enterobactin biosynthesis cistrons *entB* and *entA*. J. Bacteriol. **171**:784–790.
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. **Nichols, B. P., M. van Cleemput, and C. Yanofsky.** 1981. Nucleotide sequence of *Escherichia coli trpE*. J. Mol. Biol. **146**:45–54.
21. **Norrander, J., T. Kempe, and J. Messing.** 1983. Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis. Gene **26**:101–106.
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. **Pettis, G. S., T. J. Brickman, and M. A. McIntosh.** 1988. Transcriptional mapping and nucleotide sequence of the *Escherichia coli fepA-fes* enterobactin region. J. Biol. Chem. **263**: 18857–18863.
24. **Pettis, G. S., and M. A. McIntosh.** 1987. Molecular characterization of the *Escherichia coli* enterobactin cistron *entF* and coupled expression of *entF* and the *fes* gene. J. Bacteriol. **169**:4154–4162.
25. **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.
26. **Queen, C., and L. J. Korn.** 1984. A comprehensive sequence analysis program for the IBM personal computer. Nucleic Acids Res. **12**:581–599.
27. **Reznikoff, W. S., D. A. Siegele, D. W. Cowing, and C. A. Gross.** 1985. The regulation of transcription initiation in bacteria. Annu. Rev. Genet. **19**:355–387.
28. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
29. **Schwyn, B., and J. B. Neilands.** 1987. Universal chemical assay for the detection and determination of siderophores. Anal. Biochem. **160**:47–56.
30. **Shine, J., and L. Dalgarno.** 1974. The 3′-terminal sequence of *Escherichia coli* 16S RNA: complementarity to nonsense triplets and ribosome binding sites. Proc. Natl. Acad. Sci. USA **71**:1342–1346.
31. **Soberon, X., L. Covarrubias, and F. Bolivar.** 1980. Construction and characterization of new cloning vehicles. IV. Deletion derivatives of pBR322 and pBR325. Gene **9**:287–305.
32. **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.