The **hisT-purF** Region of the *Escherichia coli* K-12 Chromosome

**IDENTIFICATION OF ADDITIONAL GENES OF THE **hisT** AND **purF** OPERONS**

(Received for publication, November 4, 1986)

Michael L. Nonešt, Christopher C. Marvel, and Dean R. Tolan

From the Department of Biochemistry, University of California at Berkeley, Berkeley, California 94720

A 9.7-kilobase pair segment of the *Escherichia coli* chromosome spanning the **hisT** and **purF** loci has been characterized. Six genes were identified in this region by complete DNA sequence analysis, *in vivo* expression in maxicells, and RNA transcript analysis. S1 nuclease analysis has demonstrated that some of these genes are part of the **hisT** or **purF** operons. Two of the newly identified genes, **dedA** and **dedB**, were localized immediately downstream of **hisT** in the **hisT** operon. Two other genes, denoted **dedC** and **dedD**, have been localized between the **hisT** and **purF** operons. The other two genes, **dedE** and **dedF** flank the **purF** gene. **dedE** has been previously described as the first gene in the **purF** operon (Makaroff, C. A., and Zalkin, H. (1985) *J. Biol. Chem.* 260, 10378-10387). **dedF** was localized downstream from **purF** and is part of the **purF** operon. In addition, **dedF** is homologous to the **ubiX** gene of *Salmonella typhimurium*. Adjacent to **dedF** is the *E. coli* homologue of the *S. typhimurium* argT locus encoding the lysine/arginine/ornithine-binding protein.

The **hisT** gene has been cloned and demonstrated to be a component of a multigene operon (8). Sequence and genetic analysis of a cloned DNA segment containing the **hisT** gene demonstrated that **hisT** and a gene of unknown function (denoted **usg**) are translationally coupled (9). S1 nuclease mapping indicated the possible existence of additional genes in the **hisT** operon (9). Tightly linked to **hisT** is the **purF** locus, which codes for glutamine-3-phosphoribosylamine:pyrophosphate phosphoribosyltransferase (EC 2.4.2.14), the first enzyme in *de novo* purine biosynthesis (10). This gene, like **hisT**, is a component of a polycistronic operon (11).

We have examined the region between the **hisT** and **purF** genes in order to characterize additional genes of the **hisT** operon. This segment of the chromosome has been characterized by DNA sequence analysis, *in vivo* expression in maxicells, and S1 nuclease analysis. We have defined two additional genes for the **hisT** operon and one for the **purF** operon. These two operons are separated by two other genes **dedC** and **dedD**. Finally, this study identified the *E. coli* homologs for the *Salmonella typhimurium* **ubiX** and **argT** loci.

The DNA analyzed in these studies was obtained from the previously described plasmid p210. This plasmid consists of a 9.7-kb *E. coli* HindIII DNA fragment (derived from the Carbon-Clarke plasmid pLC28-44) inserted into pBR322. The pLC28-44 plasmid complements **hisT**, **hisJ**, and **purF** phenotypes (8, 12). The sequence of the first 2.3 kb of 210 which contains the **hisT** gene in a multigene operon (9) and an additional 2.7 kilobases containing the **purF** gene in a multigene operon (10-11) has been reported. In this report we describe the characterization of the remaining portion of the p210 plasmid.

**EXPERIMENTAL PROCEDURES**

**Materials**

Restriction endonucleases were purchased from Bethesda Research Laboratories (BRL) or New England Biolabs. T4 DNA ligase was from New England Biolabs. DNA polymerase I, large fragment, was from Boehringer Mannheim. [α-32P]dCTP (400 Ci/mol) was purchased from Amersham Corp., and [14C]amino acid mixture (55 mCi/mmol) was purchased from ICN Pharmaceuticals.

**Bacterial Strains**

All strains used in these studies were derived from *E. coli* K-12. M13 cloning used the strain JM103 (13). Maxicell assays used CSR603 (14). *E. coli* strain TX158 (15) containing lac genes inserted into the **purF** gene and TX140, the parental strain, were gifts of Dr. John Smith (Department of Biochemistry, Louisiana State University, Shreveport). *E. coli* W3110 and a derivative NU399, with a

*The abbreviations used are: kb, kilobase pair(s); ×1 Denhardt's, 0.01% each of polyvinylpyrrolidone, bovine serum albumin, and Ficoll; ×1 SSC, 0.15 M sodium citrate and 0.15 M sodium chloride, pH 7.0; bp, base pair(s).*
The histT-purF Region

Plasmid DNA was isolated from E. coli JA200 transformed with plasmid pLC28-44. DNA (0.3 μg) was digested with 20 units of NaeI for 24 h at 37 °C. The digestion products were separated on a 0.6% Tris-phosphate agarose gel and blotted onto nitrocellulose. The blots were soaked directly in dimethylsulfoxide for 30 min before drying and radioautography. The dried blots were at -80 °C with an intensifying screen.

Southern Analysis of pLC28-44

The construction of plasmid pLC28-44 is described by Clarke and Carbon (17) and construction of p210 is described by Marcel et al. (8). Standard recombinant DNA procedures were used as described by Maniatis et al. (20). Each clone is discussed below. The locations of restriction endonuclease sites used for the constructions are shown in Fig. 3. Verification of the constructions was performed by restriction enzyme digestions and/or sequence analysis.

pKK—The 2396-bp KpnI fragment of 9210 was inserted into pUC18. The insert was oriented such that dedB transcription is opposite lacZ’ transcription.

pEE—The 2636-bp EcoRI fragment of p210 was inserted into pUC8. The orientation was the same as pKK.

pEN—The clone, pEE, was digested with PstI and NruI and the 2626-bp fragment inserted into pBR322, which had been digested with EcoRI, made blunt by incubation with the large fragment of DNA polymerase 1, then digested with PstI. This inactivated the bla gene and allowed selection by tetracycline resistance.

pP1065, pP1180, and pP1508—p210 was digested with PvuII and the ends made blunt as above. HindIII linkers (BRL/Gibco) were added and the fragments cloned into pUC8. The resultant clones were screened by their insert size, and their orientations were confirmed by restriction endonuclease digestions. The clone pP1180 was oriented in pEE, and pKK, pP1508 was oriented in the opposite direction. pP1065 had the orientation such that lacZ’ transcription was similar to that of dedE.

pCS—A 2525-bp ClaI to Sall fragment of p210 was inserted into pBR322 at the homologous sites.

pS3—p210 was digested with SstII, diluted, and religated. The resultant plasmid had a deletion from 2967 to 7296.

pSF—pP1508 was digested with Sall which deleted a fragment from the Sall site of the vector to the Sall site at 4571. This created an in-frame fusion between the dedC and lacZ’ genes.

pS5500 and pS9000—p210 was partially digested with Sall, diluted, and religated. Clones were selected according to size, and restriction mapping confirmed the inserts. pS5500 contained a deletion from 622 bp in the vector’s tetracycline resistance gene to the Sall site at position 7897 of p210. Sall was similar to S5500 except the deletion extended only to the Sall site at position 4571.

pBgh—p210 was digested with BglII and BamHI, diluted, and religated. The resulting clone contained a deletion of 546 bp of the tetracycline resistance gene to the BglII site at position 8953 bp in the p210 insert.

pStuStuPst—p210 was digested with Stul, and either PstI or Sall. The fragments from 4571 to 6059 (Sall to Stul) and from 6059 to 9687 (Stul to PstI) were purified and cloned into Sall-Msal or PstI-Sall digested pUC9, respectively. The orientation of pStuStu was such that lacZ’ transcription was the same direction as dedD, and pStuPst was oriented such that lacZ’ transcription was opposite purF.

pElb—p210 was digested with EcoRI and BglII, and the fragment from 3857 to 7677 was inserted into BamHI-EcoRI-restricted pUC8.

pBbgBg—BglII and PstI digested p210 was ligated into BamHI-digested pUC8 and the clones which contained the BglII fragment from position 7677 to position 8953 were sequenced by the method of Maxam and Gilbert (25).

Plasmid-directed Protein Synthesis in Maxicells

Maxicells were prepared and analyzed according to Sancar et al. (14) with two modifications. Cells were labeled with a [35S]amino acid mixture instead of [3H]methionine and 0.1% phenylmethylsulfonyl fluoride was added to the cell lysis buffer to minimize degradation of labeled proteins. The protein products were electrophoresed on sodium dodecyl sulfate-polyacrylamide gels with a gradient of 12-16% in acrylamide. Gels were first stained with Coomassie Blue or soaked directly in dimethylsulfoxide for 30 min, in 20% 2,5-diphenyloxazole followed by dimethyl sulfoxide for 45 min and water for 60 min before drying and radioautography.

S1 Nuclease Mapping

DNA from M13 single-stranded recombinant clones served as templates for the synthesis of radioactively labeled second strands complementary to transcripts from this region of the E. coli genome. The oligonucleotide primer used for sequence determination was DNA was sequenced by the method of Sanger et al. (21), using M13mp8, M13mp9, M13mp18, and M13mp19 (22,23). Regions with dG-dC compression were sequenced by the substitution of ddGTP with 7-deaza-ddGTP (24) in the second-strand synthesis. Some regions were sequenced by the method of Maxam and Gilbert (25).

FIG. 1. Orientation of p210 to the E. coli chromosome. Panel A, the order of the hisT, purF, and his’ loci and their chromosome position in min. Map positions were derived from the linkage map of E. coli (30). Reading left to right is counter-clockwise on the E. coli chromosome. Panel B, plasmid, pLC28-44. The gray area is the vector colicin E1 and the dotted area is that portion subcloned into pBR322 to create p210. E, EcoRI, H, HindIII restriction sites were determined by Marcel et al. (8). The arrows inside the circle show the locations of the mapped operons as described in the text. Panel C, hybridization of probes to NaeI-digested pLC28-44 DNA blotted onto nitrocellulose. The digestion yielded two major and one minor ethidium-bromide-staining fragments of 10, 9, and 8.5 kb, respectively. The smallest band is likely due to digestion at a refractory NaeI site external to the dotted region (panel B). The dotted region corresponded to the region shown in panel B. Hybridization was for 24 h at 65 °C with 1 x 106 cpm/ml of probe DNA. Lane 2, as in lane 1 except probe was a 730-bp EcoRI fragment shown as part of the hisT operon in panel B. Lane 3, as in lane 1 except probe was a HindIII fragment isolated from the S. typhimurium hisJ gene. Hybridization was done at 60 °C.
The hisT-purF Region

Fig. 2. Restriction map of the ϕ210 E. coli DNA insert. Below the restriction map are boxes delineating the genes identified in this region. The smaller boxes indicate genes previously characterized. Solid bars depict the regions constructed as subclones of ϕ210. The dotted line in pSA denotes the region deleted. The names of the clones are derived from the restriction enzyme sites at the termini. Where some clones used the same sites, the bp size of the insert is used. The arrows at the bottom depict the direction and extent of sequencing. Where the arrows are connected, the clones were composed of several fragments in one M13 clone. The shaded arrow denotes the sequence determined by the method of Maxam and Gilbert (25). The gray bar at the bottom denotes the area sequenced by Arps et al. (9), the rightward or leftward slanting cross-hatched bars denote the regions sequenced by Makaroff and Zalkin (11) or Tao et al. (10), respectively. The restriction sites at the top are denoted by: H, HindIII; E, EcoRI; C, ClaI; K, KpnI; P, PvuII; Sst, SstII; N, NruI; U, Stul; S, SalI; Nae, Nael; B, BglII; Pst, PstI; V, PvuII.

Analysis and manipulation of sequence data were performed on a DEC VAX-780 and VAX-750 computers using sequence analysis software developed by Dr. Hugo M. Martinez (Biomatics Computation Laboratory, Department of Biochemistry and Biophysics, University of California at San Francisco, San Francisco). Homology searches used the program "dbalign" which was run against GenBank, release 35.0, 1 August 1985.

Additional analysis was performed using the ANALSEQ package (Whitehead Institute for Biomedical Research, Cambridge MA) for running the positional base preference program of Staden (29).

RESULTS

Orientation of the ϕ210 Insert on the E. coli Chromosome

In E. coli, purF and hisT have been genetically mapped near hisJ in the order of HisJ-purF-hisT (30). In addition, phage lambda integrations had been used to determine a clockwise direction of transcription for the purF gene (15). However, our sequence analysis demonstrated that purF was transcribed away from hisT, and thus counter-clockwise using the present gene order (Fig. 1, panel A). This contradiction could be resolved by assuming a revised gene order of hisJ-hisT-purF which would allow clockwise transcription of purF, away from hisT. To resolve this issue a restriction digestion of pLC28-44 DNA was analyzed with hybridization probes for hisT and hisJ. Restriction digestion of pLC28-44 with NaeI yielded three fragments. The smaller two fragments must have derived from the same region of the plasmid as they both hybridized to a probe for hisJ from S. typhimurium, and the combined sizes of all three was greater than the size of ϕ210. This doublet is likely due to a site which is refractory to Nael digestion, such as is found in pBR322. The other fragment of 10 kb hybridized to the hisT probe (Fig. 1, panels B and C). The gene probes clearly hybridized to two separate fragments, and since a known Nael site is present in purF (10), the two genes must be separated by purF. Therefore, hisJ does not map between purF and hisT and the gene order is hisJ-purF-hisT, which forces the conclusion that purF is transcribed counter-clockwise on the E. coli map.

Sequencing of ϕ210

The Sanger dideoxy DNA-sequencing method was used with M13 vectors to sequence the 4.5 kb separating purF and hisT and the 1.4 kb located distal to the purF gene. This completes the entire sequence of the 9689 bp E. coli DNA fragment contained in ϕ210. The sequencing strategy is shown at the bottom of Fig. 2 with arrows indicating the orientation and size of the individual sequenced clones. The DNA sequence was determined bidirectionally for 80% of the sequence. The regions of published sequence are indicated by bars at the bottom of the figure. Fig. 3 has the DNA sequence and derived amino acid sequences of the identified coding frames. dG·dC-rich regions were sometimes ambiguous due to compression. These were resolved using dITP and d7-dGTP (24) instead of dGTP during second-strand synthesis. The sequence determined on both sides of purF overlapped the published data (10, 11).

Analysis of Coding Probabilities of Open Reading Frames Identified by DNA Sequencing

Demonstration that the open reading frames in the DNA sequence were consistent with protein-coding regions in E. coli was done by computer analysis. The first analysis was a positional base preference method described by Staden (29). This program identifies coding sequence by amino acid composition with no weighting for codon preference. Results of this analysis are shown in Fig. 4. Nine peaks representing
FIG. 3. The sequence of the *E. coli* chromosome between hisT and argT. The start of each protein is denoted by an arrow preceded by the name of the gene (also denoted at the left margin). The *"* amino acid for each codon is given in one letter code. Restriction sites important for the interpretation of the maxicell denoted by an arrow preceded by the name of the gene (also denoted at the left margin). The amino acid for each below their underscored sequence. The palindromic regions which may function as p-independent terminators (33) as dashed lines. The vertical arrow between the dedD and dedE genes denotes the start of transcription for the purA operon. The start of transcription for the hisT operon (11). The bases between two adjacent genes denotes a repetitive extragenic palindromic sequence (11, 37) identified by a homology search in GenBank. The regions homologous to *ubi* and *argT* are denoted by brackets. The numbers at the right correspond to the nucleotide positions of the entire *"* insert beginning at the * HindIII site in the hisT operon of *"*. The sequence from position 1 to 2164 and between 6897 and 8246 has been previously published (9-11) and is omitted from the depicted sequence.
were probable protein-coding regions. Verification that these open reading frames identified by DNA sequencing were similar to the codon usage possible for E. coli revealed only one probable region using the codon preferences found in the reverse complement, panel B, and previously; this implicated a truncation of the dedA promoter.

Codon usage was analyzed in the open reading frames identified in the sequence using the codon preferences found in E. coli. These data are presented in Table 1. The codon usage was expressed as a percentage of the most optimum codon usage possible for E. coli (31). All of the transcripts encoded on \( \psi 210 \) had values from 0.6 to 0.76. These values were similar to the lacY and trpB genes but less than the values for highly expressed genes. This analysis demonstrated that the open reading frames identified by DNA sequencing were probable protein-coding regions. Verification that these DNA regions actually code for proteins was done by analysis in a maxicell protein expression system (14).

Location of Nine Genes Expressed in Maxicells in \( \psi 210 \)

Maxicell assays revealed at least nine proteins encoded on \( \psi 210 \) ranging in molecular weight from 19 to 56 kDa (Fig. 5, panels A–C, lane 1). Of these nine, four have been described previously; hisT, usg (8, 9), purF (10) and one upstream from purF (11). The genes characterized in this report have been designated dedA through dedF for downstream E. coli DNA (from hisT).

Extensive subcloning of \( \psi 210 \) followed by maxicell analysis confirmed the location, size, and orientation of each of the ded genes. The subclone designations used refer to a DNA fragment cloned into a plasmid vector and transformed into the maxicell strain CSR603. The insert DNA of plasmids used in these experiments are depicted by the solid bars in Fig. 2. dedA—The position of the dedA gene was determined by its presence in the maxicell assays of pEN and pS\( \Delta \) shown in Fig. 5, panel A, lanes 5 and 7. The assay of pS\( \Delta \) localized dedA between hisT and the SstII restriction site. The direction of transcription was indicated by the maxicell assay of pCS, which implicated a truncation of the dedA promoter.
The hisT-purF Region

Fig. 4. Predicted coding regions in \$210 using the positional base preference method of Staden (29). The 9.7-kb DNA sequence from \$210 is shown along the x axis. The probability of coding is plotted in the y direction. The method does not assume any codon preferences but instead evaluates the nonrandom use of triplets for codons in proteins as determined by an examination of the proteins in the Dayhoff database. Each reading frame (RF) is shown one on top of another. The bars at the base line for each frame represent start codons (AUG or GUG), the bars at the mid-line (50% probability) show the stop codons. The straight lines through the mid-line of each peak show a probable gene coding frame. Probabilities of coding are calculated by sliding a window of 67 codons along the sequence one codon at a time. For each position of the window, the codons in each of three reading frames of the DNA are compared with the standard matrix from all known proteins. The corresponding probabilities of coding are calculated and plotted above one another for the four every five codons. The genes identified by the sequence are labeled in panel A, lanes 1–5. dedB—The position of the dedB protein is shown in panel A, lanes 5–7. dedB protein was present in maxicell assays of pCS and absent in maxicell assays of pSÁ. This located the gene between the SstII site and the SalI site. The assay of pEN revealed a truncation product of 27 kDa in place of the native dedB product. The direction of transcription for dedB gene was toward purF.

dedC—The position of dedC was defined by the maxicell assays of pP1508+ and pSP shown in panel B, lanes 3 and 4. The assay of pP1508+ contained a full-sized dedC gene product. This and dedCs absence in assays of pSÁ located the gene between dedB and the PvuI site at position 5392. Assays of pSP gave a smaller protein of 37 kDa. This provided evidence for the direction of transcription, as an in-frame fusion was created between the lacZ' reading frame of pUC8 and the dedC-coding region which accounted for the dedC* protein product.

dedD—The position of dedD was determined by maxicell assaying pS9000 and pEBg as shown in panel D, lanes 1 and 2. The presence of a dedD gene product from both of these clones localized the gene between the SalI site at position 4572 and the BglII site at position 7875. The presence of dedD in pEBg and the presence of purF in pS9000, combined with the absence of dedD in pSÁ (lane 3), confined the gene between dedC and purF. The analysis of two clones which split this region, pSStu and pStuPst, are shown in panel D, lanes 4 and 5. The analysis of the pStuPst clone revealed the dedD product, whereas that of pSStuPst revealed the position of the dedE, purF, and dedD, but no products originating from dedD. This evidence located dedD between dedC and dedE. The results from the pStuPst clone showed that the genes of the purF operon were transcribed independently of dedD. Evidence for the direction of transcription of dedD toward purF was determined due to the singular open reading frame present in the DNA sequence (Fig. 4). The data do not allow an unambiguous prediction of the translational start site for the dedD protein. The open reading frame between the end of the dedC gene and the StuI site did not contain a AUG codon until position 5519. This AUG is not thought to be used, as the clone pP1065 did not express a protein from the transcript initiated from the lacZ' promoter of this plasmid. There exist several GTG codons in the dedD reading frame near the end of the dedC open reading frame. Two have conserved Shine-Dalgarno ribosome-binding sites (35) upstream from a GTG codon, although the furthest one would cause an overlap of four codons of the dedC gene (See Fig. 3).

dedE—The position of dedE was defined as shown in panel D, lanes 1–6. The absence of dedD and the presence of dedE in the maxicell assay of the pSStuPst clone (lane 5) localized the dedE gene between dedD and purF. The assay of pP1065 revealed a truncation product of 12 kDa (dedE*), which was consistent with a truncation of the dedE gene. This gene was identical to the first protein in the purF operon described previously (11).

dedF—The location of dedF was determined by maxicell assays of the clones shown in panel C, lanes 2–5. Expression of dedF was independent of purF as shown by assays of the pSS5500 clone. This result localized the gene between the SalI site at position 7895 and the end of the \$210 insert. The gene product was not expressed by the pBGH clone (lane 5), which eliminated the possibility that the dedF product represented the amino terminus of a larger gene which extended outside of \$210. The location and the direction of transcription were defined by assays of the clones pBGH* and pBGH1 (lanes 3 and 4). Proteins of 25 kDa and a 23.5 kDa could be explained as fusion products between the dedF reading frame and the lacZ' polypeptide or an open reading frame on the opposite strand of the lacZ', respectively (dedF*). In this analysis the

| Gene | Frequency | Gene | Frequency |
|------|-----------|------|-----------|
| usg  | 0.76      | recA | 0.85      |
| hisT| 0.64      | trpB | 0.70      |
| dedA| 0.60      | lacY | 0.61      |
| dedB| 0.74      | araC | 0.54      |
| dedC| 0.65      |      |           |
| dedD| 0.60      |      |           |
| dedE| 0.62      |      |           |
| purF| 0.70      |      |           |
| dedF| 0.60      |      |           |

2. The presence of a dedD gene product from both of these clones localized the gene between the SalI site at position 4572 and the BglII site at position 7875. The presence of dedC in pEBg and the presence of purF in pS9000, combined with the absence of dedD in pSÁ (lane 3), confined the gene between dedC and purF. The analysis of two clones which split this region, pSStu and pStuPst, are shown in panel D, lanes 4 and 5. The analysis of the pStuPst clone revealed the dedD product, whereas that of pSStuPst revealed the position of the dedE, purF, and dedD, but no products originating from dedD. This evidence located dedD between dedC and dedE. The results from the pStuPst clone showed that the genes of the purF operon were transcribed independently of dedD. Evidence for the direction of transcription of dedD toward purF was determined due to the singular open reading frame present in the DNA sequence (Fig. 4). The data do not allow an unambiguous prediction of the translational start site for the dedD protein. The open reading frame between the end of the dedC gene and the StuI site did not contain a AUG codon until position 5519. This AUG is not thought to be used, as the clone pP1065 did not express a protein from the transcript initiated from the lacZ' promoter of this plasmid. There exist several GTG codons in the dedD reading frame near the end of the dedC open reading frame. Two have conserved Shine-Dalgarno ribosome-binding sites (35) upstream from a GTG codon, although the furthest one would cause an overlap of four codons of the dedC gene (See Fig. 3).
The hisT-purF Region

**Fig. 5.** Maxicell plasmid directed in vivo protein synthesis. The products were analyzed by electrophoresis on 12-16% sodium dodecyl sulfate-polyacrylamide gels. Marker proteins of known molecular weight were co-electrophoresed, and their positions are given at the right of panel A. The arrows to the left in each panel denote the positions of the major protein products seen in maxicell experiments using the entire ψ210 plasmid (lane 1, all panels). The products produced from selection markers on the vectors were either tet or bla gene products. The β-lactamase products can appear as one, two, or three different molecular species; 30 and 31 kDa mature products and a 32.5 kDa precursor. The tetracycline gene product is present as a selection marker.

**Panel A** shows in vivo protein synthesis. The positions of the major protein products are indicated with asterisks (*). The gel was radioautographed for 5 days. The major protein products are denoted by the presence of a common marker protein of known molecular weight. The positions of the major protein products in each lane are indicated by arrows.

**Panel B** shows the localization of dedA and dedD. Lane 1, ψ210; lane 2, pUC8; lane 3, pBR322 selected with ampicillin; lane 4, the maxicell strain, CSR603 without a plasmid; lane 5, pCS; lane 6, pCSS; lane 7, pEN.

This gel was radioautographed for 6 days except lane 7 which was for 2 weeks. Panel B, maxicell assays showing the localization of dedC. Lane 1, ψ210; lane 2, pUC8; lane 3, pP1065. Radioautography for 10 days. The asterisk (*) denotes truncated protein products and the § denotes fusion gene products.

The direction of transcription was the same as purF.

Analysis of the DNA sequence of ψ210 predicted open reading frames coding for the ded proteins A–F, of molecular mass 24,500, 33,400, 45,500, 23,000, 17,900, and 20,800 daltons, respectively. The apparent molecular weights of the proteins based on sodium dodecyl sulfate-polyacrylamide gel electrophoresis were 19,500, 33,000, 49,000, 35,500, 16,500, and 21,000, respectively. Except for dedA and dedD the size of each open reading frame was within 10% agreement with its apparent molecular weight. Amino acid compositions may account for these differences, for instance, dedD is proline rich (14%). These effects are similar to what has been described for the usg protein of the hisT operon (8). The deduced amino acid sequence of each protein is found in Fig. 3 above the DNA sequence of their respective genes.

Arps et al. (9) have shown that usg and hisT genes are transcribed on a single RNA. In order to test if the dedA gene is coordinately transcribed with hisT, a single-stranded DNA clone bridging the hisT-dedA reading frames (CK-3) was used in S1 nuclease protection experiments (Fig. 6). A similar experiment was performed in order to test if dedB is coordinately transcribed with dedA using a clone bridging dedA and dedB (3a). Full length protected fragments were found in both cases when RNA from wild type E. coli was used. The apparent sizes of these full length protected fragments are shown and correspond well with the expected sizes. However, RNA from a polar hisT mutant (strain NU399) failed to protect the full length DNA probes in both experiments. This demonstrated that transcripts exist which, 1) cross the intergenic regions between hisT-dedA and dedA-dedB, and 2) that these transcripts likely initiate upstream from the hisT polar mutation. In an analogous experiment, the full length of a single-stranded clone, BBF-342, which bridges purF-dedF reading frames was protected by RNA from wild type E. coli but not by RNA from the purF polar mutant TX158. This demonstrated that transcripts exist which cross the purF-dedF intergenic region and which likely initiate upstream from the purF polar mutation.

**Sequence Comparison to the GenBank Database**

Identification of sequence homologies between ψ210 and published data was accomplished by a homology search of the GenBank database. Several sections of ψ210 were homologous to GenBank files.
ubiX—The amino acid sequence of dedF was 70% similar to the sequence identified in *S. typhimurium* as ubiX. 2

argT—Immediately following dedF and extending to the end of the *ψ210* insert was an open reading frame of 146 amino acids which was 89% similar to the argT locus of *S. typhimurium*, the gene for the lysine/arginine/ornithine-binding protein (32).

**DISCUSSION**

The description of six genes downstream from the hisT operon, and evidence for the coordinate transcription of three of these genes with the hisT and purF operons, has been accomplished using three independent criteria. First, the DNA sequence data predicted coding regions with codon usage similar to that of moderately expressed *E. coli* proteins. In addition computer analysis identified likely protein-coding reading frames on *ψ210*. Second, the proteins expressed in a maxicell system were consistent with each gene’s orientation, position, and size. Third, S1 nuclease mapping identified additional genes which are coordinately transcribed with either hisT or purF.

The hisT gene has been previously determined to be part of an operon. In addition, previous S1 nuclease analysis had demonstrated that transcription continues past the hisT gene (9). Furthermore, in this study, S1 nuclease mapping of transcripts from wild type and a hisT polar mutant has demonstrated that dedA and dedB are additional components of the hisT operon. Although the operon may extend further downstream from dedB, a consensus p-independent terminator sequence (33) is located immediately following the coding sequence for dedB (40354057). This sequence could form an 11-base pair GC-rich stem followed by 4 thymidine residues and may be the terminator for the hisT operon, however, this has not been experimentally demonstrated. Taken together the hisT operon consists of at least four cistrons of which only one encodes a protein of known function.

Immediately downstream of dedB two genes have been identified which are termed dedC and dedD. These genes may have been previously mentioned in earlier studies. Makaroff and Zalkin (11) reported that an upstream transcript terminated near the purF operon promoter, and an uncharacterized protein (47 kDa) was produced from a 1.5-kb fragment of DNA that extends 5’ upstream of purF. The immediate upstream transcript likely corresponds to the dedD gene, and the 47 kDa protein is probably the dedC gene product. In addition, Bogner et al. (34) have reported the cloning of the folC gene. This gene is present on the plasmid pLC28-44, has a restriction map which indicates that it lies in the region identified here as dedC, and the purified folC product (molecular mass approximately 47 kDa), polyglutamate synthet-
tase-dihydrofolate synthetase, closely corresponds to the molecular mass of the dedC protein (45.5 kDa). It is possible that dedC and folC are equivalent loci.

Immediately downstream from dedD are the genes comprising the purF operon. The first two genes dedE and purF have been described previously (11). Downstream from the purF gene is dedE which has been identified as a component of the purF operon by S1 nuclease mapping and whose transcription is affected by a purF polar mutation. Moreover, in several maxicell clones dedE was transcribed independently of dedE, purF, or known vector promoters, which may indicate alternate expression independent of the operon promoter. The start of transcription for this operon has been determined (11) and is shown in Fig. 3 (position 6233). In addition, the sequence after dedE contains a consensus p-independent termination structure (8991–9013) which may be a terminator of the purF operon. However, this proposed terminator has not been experimentally defined. The purF operon appears to consist of the three genes, dedE, purF and dedF.

Previous genetic studies in this region are limited. Hong and Ames (36) have mapped five lethal temperature-sensitive mutations between hisT and purF in S. typhimurium. Some of these mutations map nearer to hisT than to hisT and could be localized in dedE. This is supported by the observation of Makaroff and Zalkin (11) that no polar purF mutations could be mapped to the dedE gene in E. coli. It is also possible that some of these mutations map to other ded genes. Further analysis of these previously isolated mutants and mutational analysis of this region will be required to define the function of the ded genes.

Acknowledgments—We would like to thank Dr. Edward E. Penhoet in whose laboratory the majority of this work was done. Our appreciation is given to Dr. G. F-L. Ames at the University of California, Berkeley for allowing us access to her unpublished data and for critical reading of the manuscript. Thanks to Dr. H. Kammen and Dr. M. E. Winkler for critical reading of the manuscript and G. Mardon for helpful discussions. We would like to express our appreciation for the technical assistance of Kelly DeSelms and Patty Holman and to Michael Stern for making available a probe for hisJ.

REFERENCES
1. Chang, G. W., Roth, J. R., and Ames, B. N. (1971) J. Bacteriol. 108, 410–414
2. Parker, J. (1982) Mol. Gen. Genet. 187, 405–409
3. Singer, C. E., Smith, G. R., Cortese, R., and Ames, B. N. (1972) Nature New Biology 238, 72–74
4. Cortese, R., Kammen, H. O., Spengler, S. J., and Ames, B. N. (1974) J. Biol. Chem. 249, 1103–1108
5. Kammen, H. O., and Spengler, S. J. (1975) Fed. Proc. 34, 517
6. Turnbough, C. L., Neill, R. J., Landsberg, R., and Ames, B. N. (1979) J. Biol. Chem. 254, 5111–5119
7. Rosenfeld, S. A., and Brenchley, J. E. (1980) J. Bacteriol. 143, 801–808
8. Marvel, C. C., Arps, P. J., Rubin, B. C., Kammen, H. O., Penhoet, E. E., and Winkler, M. E. (1986) J. Bacteriol. 161, 60–71
9. Arps, P. J., Marvel, C. C., Rubin, B. C., Tolan, D. R., Penhoet, E. E., and Winkler, M. E. (1985) Nucleic Acids Res. 13, 5297–5315
10. Tso, J. Y., Zalkin, H., van Cleemput, M., Yanofsky, C., and Smith, J. M. (1982) J. Biol. Chem. 257, 3525–3531
11. Makaroff, C. A., and Zalkin, H. (1985) J. Biol. Chem. 260, 10378–10387
12. Ardeshir, F. (1981) Ph. D. thesis, Department of Biochemistry, University of California at Berkeley
13. Messing, J., Crea, R., and Seeburg, P. H. (1981) Nucleic Acids Res. 9, 309–321
14. Sancar, A., Hack, A. M., and Rupp, W. D. (1979) J. Bacteriol. 137, 692–693
15. Smith, J. M., and Gots, J. S. (1980) J. Bacteriol. 143, 1156–1164
16. Vieira, J., and Messing, J. (1982) Gene 19, 259–268
17. Clarke, L., and Carbon, J. (1976) Cell 9, 91–99
18. Davis, R. W., Botstein, D., and Roth, J. R. (1980) Advanced Bacterial Genetics, pp. 202–203, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
19. Rigby, P. W. J., Dieckmann, M., Rhodes, C., and Berg, P. (1977) J. Mol. Biol. 113, 257–251
20. Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual, pp. 3–186, 390–402, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
21. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U. S. A. 74, 5463–5467
22. Messing, J., and Vieira, J. (1982) Gene 19, 269–276
23. Norrander, J., Kempe, T., and Messing, J. (1983) Gene 26, 101–106
24. Barr, P. J., Thayer, R. M., Laybourn, P., Wejerian, R. C., Seela, C., and Tolan, D. R. (1986) Biotechniques 4, 428–432
25. Maxam, A. M., and Gilbert, W. (1980) Methods Enzymol. 65, 499–560
26. Burgess, D. G., and Penhoet, E. E. (1985) J. Biol. Chem. 260, 4604–4614
27. Buchman, A. R., Fromm, M., and Berg, P. (1984) Mol. Cell. Biol. 4, 1900–1914
28. Salsen, W., Gesteland, R. F., and Bolle, A. (1967) Nature 215, 598–591
29. Staden, R. (1984) Nucleic Acids Res. 12, 551–567
30. Bachmann, B. J. (1980) Microbiol. Rev. 47, 180–230
31. Ikenura, T. (1981) J. Mol. Biol. 151, 389–409
32. Higgins, C. F., and Ames, B. F. (1981) Proc. Natl. Acad. Sci. U. S. A. 78, 6038–6042
33. Holmes, W. M., Platt, T., and Rasenbg, M. (1983) Cell 32, 1029–1032
34. Bogner, A. L., Osborne, C., Shanes, B., Signer, S. C., and Ferone, R. (1985) J. Biol. Chem. 260, 5625–5630
35. Shine, J., and Dalgarino, L. (1975) Nature 254, 34–38
36. Hong, J-S., and Ames, B. N. (1971) Proc. Natl. Acad. Sci. U. S. A. 68, 3158–3162
37. Stern, M. J., Ames, G. F-L., Smith, N. H., Robinson, E. C., and Higgins, C. F. (1984) Cell 37, 1015–1026