JOURNAL OF BACTERIOLOGY, Apr. 1986, p. 297–300
0021-9193/86/040297-04$02.00/0

Vol. 166, No. 1

Chromosomal Location and Nucleotide Sequence of the *Escherichia coli dapA* Gene

FRANCOIS RICHAUD,* CATHERINE RICHAUD, PASCAL RATET, AND JEAN-CLAUDE PATTE†

Institut de Microbiologie, Université Paris Sud, 91405 Orsay Cedex, France

Received 25 November 1985/Accepted 24 January 1986

In *Escherichia coli*, the first enzyme of the diaminopimelate and lysine pathway is dihydrodipicolinate synthetase, which is feedback-inhibited by lysine and encoded by the *dapA* gene. The location of the *dapA* gene on the bacterial chromosome has been determined accurately with respect to the neighboring *purC* and *dapE* genes. The complete nucleotide sequence and the transcriptional start of the *dapA* gene were determined. The results show that *dapA* consists of a single cistron encoding a 292-amino acid polypeptide of 31,372 daltons.

The *dapA* gene encodes dihydrodipicolinate synthetase (DHDPS) (10), the first enzyme of the diaminopimelate and lysine pathway (23). It has been mapped to 53 min on the *Escherichia coli* chromosome (1). DHDPS activity is sensitive to lysine inhibition (28) as is aspartokinase III (26), the first enzyme of the biosynthetic pathway leading from aspartate to diaminopimelate, lysine, methionine, threonine, and isoleucine. The regulation of expression of several genes of the lysine regulon (19), including *lysC* (26), *asd* (4, 10), *dapB* (13), *dapD* (21), *dapE* (10), and *lysA* (20), has been studied either by directly measuring their gene products or by the use of *lacZ* or *galK* fusions. In contrast to all these genes, whose levels of expression depend on the lysine pool, the synthesis of DHDPS does not appear to be subject to any regulatory control (7, 23). Moreover, it has been shown that DHDPS must catalyze the rate-limiting step in lysine biosynthesis after aspartokinase III (11).

A thorough knowledge of this step is necessary for the understanding of the whole lysine pathway. Consequently, the location of the *dapA* gene on the *E. coli* chromosome has been precisely determined in relation to the neighboring genes. The *dapA* gene has been identified in an *E. coli* λ library, purified, and sequenced, and its transcriptional start has been determined.

MATERIALS AND METHODS

Media, strains, and plasmids. Bacterial strains and plasmids are listed in Table 1. Growth conditions were as described previously (21).

DNA manipulations. Plasmid purification; DNA fragment isolation, restriction, and ligation; plasmid transformation; DNA sequencing; and determination of transcriptional start were performed as described previously (21).

Enzyme assays. DHDPS was measured by the *O*-aminobenzaldehyde assay of Yugari and Gilvarg (28) as modified by Butour et al. (7). Samples of products were taken after 30, 45, and 60 min of incubation at two different protein concentrations. Units are given in increments of optical density (10^3) at 540 nm per minute per milligram of protein. Protein concentrations were determined by the biuret method. β-Lactamase assays were performed by the method of Novick (17).

RESULTS AND DISCUSSION

Isolation and accurate location of the *dapA* gene. Bukhari and Taylor (6) isolated three Dap mutants (AT978, AT984, and AT998) with mutations which were mapped to the *guaA* region (revised map, 53 min [1]) and identified indirectly as *dapE*, *dapA*, and *dapA*, respectively. Recently, Parker (18) located the positions of mutations in strains AT978 and AT984 close to the *purC* gene. We have previously isolated several Dap mutants (22) devoid of DHDPS activity and located close to *purC*.

The *dapA* gene has been isolated on a 2.8-kilobase (kb) *Pst*I fragment (22) subcloned from a λ bacteriophage (8) into plasmid pBR322 to give pDA1. This plasmid complements all the *dapA* mutants in our collection (22) as well as DO9M16 (5) and AT998 (6) but fails to complement the two other Dap mutants, AT978 and AT984. The latter two strains contain DHDPS activity at the wild-type level, so they are not affected in the *dapA* gene.

The restriction map of the pLC25-14 plasmid from the Clarke and Carbon library (9) carrying the *purC* gene (16) has been compared with those of the λ d*purC* phages studied by

TABLE 1. Bacterial strains and plasmids

Strain or plasmid	Relevant characteristics	Reference or source
RM4102	*araD139 ΔlacU169 rpsL thiA*	22
RDA1, 3, 4, 5, 8, 14, 25, 26	*araD139 ΔlacU169 rpsL thiA dapA*::Mu	22
DO9M16	*lys-1100 lysC1002 dapA1101*	5
AT998	*dapA16*	6
AT984	*dap-6*	6
AT978	*dapE9*	6
pDA1	Tc^r *dapA*$^+$	22
pBR322	Ap^r Tc^r	3
pUC9	Ap^r *lacZ*	27
pDA68	Tc^r *dapA* + *Bcl*I deletion of pDA1	This work
pDA2	Ap^r *dapA* + *Bst*NI fragment in pUC9	This work
pDA3	Ap^r *dapA* + *Bst*NI fragment in pUC9	This work
pDA4	Ap^r *dapA* + *Bst*NI fragment in pBR322	This work
pDA5	Ap^r *dapA* + *Bst*NI fragment in pBR322	This work

* Corresponding author.

† Present address: Laboratoire de Chimie Bacterienne, Centre National de la Recherche Scientifique BP71, 13277 Marseille Cedex 9, France.

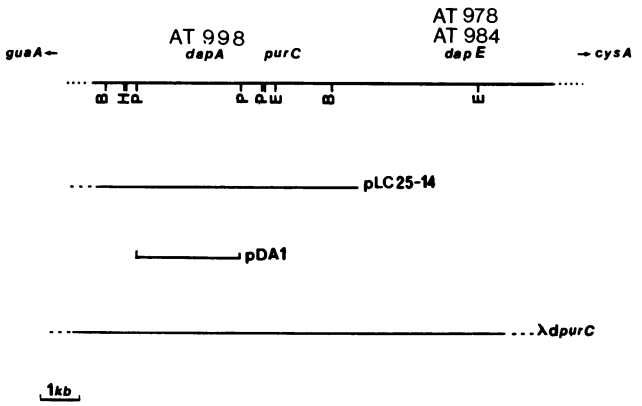

FIG. 1. Overall location of the *dap* genes in the 53-min region of the *E. coli* chromosomal map. DNA fragments used in this comparison are pLC25-14 (9) carrying the *purC* gene (16), pDA1 (22), and λ d*purC* (18). Restriction enzyme abbreviations: B, *Bam*HI; E, *Eco*RI; H, *Hin*dIII and P, *Pst*I.

Parker (18) and of our λ fragment carrying the *dapA* gene (22). Correlations between these physical maps led us to propose the revised order of *dapA*, *dapE*, and *purC* genes in the 53-min region (Fig. 1).

From plasmid pDA1 (22), the precise location of the *dapA* gene was obtained by the subcloning shown in Fig. 2. We first constructed smaller plasmids by carrying out partial *Taq*I digestions; plasmids lacking the *Taq*I fragment (*Taq*I$_4$ to *Taq*I$_5$ in Fig. 2) did not complement the *dapA* mutants and did not produce Dap$^+$ recombinants with any chromosomal *dapA* mutation. We also constructed a *Bcl*I-deleted plasmid pDA68 which complements all the *dapA* mutants. We then subcloned the 1.2-kb *Bst*NI fragment from pDA1 into the unique *Hin*cII site of plasmid pUC9 (27) in both orientations to give plasmids pDA2 and pDA3. These two plasmids complement all the *dapA* mutants we have tested, indicating that the entire *dapA* gene is localized within the 1.2-kb *Bst*NI fragment.

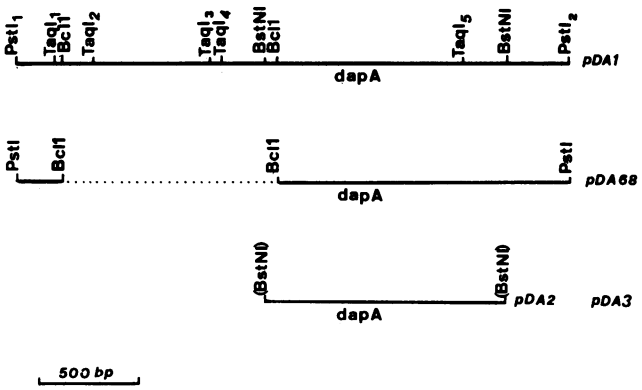

FIG. 2. Location of the *dapA* gene on cloned fragments. Plasmid pDA1 (22) has a *Pst*I fragment from a bacteriophage λ library cloned into the *Pst*I site of pBR322. Plasmid pDA68 was obtained by internal *Bcl*I deletion of plasmid pDA1. Plasmids pDA2 and pDA3 were obtained by cloning the *Bst*NI fragment from pDA1 into the unique *Hin*cII site of pUC9 (27) in both orientations. Plasmid pDA1 was digested by *Bst*NI endonuclease, and the cohesive ends were filled by DNA polymerase and then ligated to *Hin*cII-digested pUC9. bp, Base pairs.

TABLE 2. Levels of DHDPS in strain RDA8 harboring various plasmids[a]

Strain	DHDPS sp act		Plasmid copy no.[c]
	Without L-lysine	With 20 mM L-lysine[b]	
RM4102	160	20	
RDA8	5	ND[d]	
RDA8(pDA2)	7,000	700	120
RDA8(pDA3)	7,000	725	
RDA8(pDA4)[e]	3,500	400	
RDA8(pDA5)[e]	3,500	500	50

[a] Strains were grown in 63 medium supplemented with 0.4% glucose and thiamine. For RDA8, growth was in the presence of 0.5 mM *meso*-DL-diaminopimelate. For RDA8 harboring plasmids, growth was in the presence of 25 μg of ampicillin per ml.

[b] Activities are expressed as increments in optical density (10^3) at 540 nm per minute per milligram of protein (7).

[c] The plasmid copy number was determined by β-lactamase measurement by the method of Novick (17).

[d] ND, Not determined.

[e] pDA4 and pDA5 are derived from pDA2 and pDA3, respectively (Fig. 2), by cloning the *Bst*NI fragment (with the adjacent *Hin*dIII and *Bam*HI sites of the pUC9 linker) into the *Hin*dIII and *Bam*HI sites of pBR322.

Nucleotide sequence of the *dapA* gene. The nucleotide sequence of the 1.2-kb *Bst*NI fragment was determined by the method of Maxam and Gilbert (14). The strategy shown in Fig. 3 allowed complete determination on both strands. The complete sequence is given in Fig. 4. There is a large open reading frame with a potential initiating ATG and a putative ribosome-binding site, GAGG (24). This open reading frame of 292 triplets allows the synthesis of a 31,372-dalton polypeptide, which is in good agreement with the 134,000 molecular weight of the tetrameric form (23) of the native enzyme. The amino acid composition of DHDPS polypeptide deduced from the nucleotide sequencing determination is consistent with the amino acid composition determined by acid hydrolysis (23) except for methionine residues (nine deduced from the nucleotide sequence and none by amino acid analysis; this could be due to destruction of methionine sulfone during acid hydrolysis). Sixteen nucleotides downstream from the translation stop codon, another open reading frame starts, preceded by a strong ribosome-binding site, GGAG, which could be the beginning of another gene cotranscribed with *dapA*.

Determination of the transcriptional start. The 5′ end of the *dapA* mRNA has been localized by reverse transcription (25). The strategy followed is shown in Fig. 3. The transcriptional start is located at the A 24 nucleotides upstream from the translational ATG (Fig. 5). The same result is obtained by S1 nuclease protection (2) (data not shown). Upstream of this transcriptional start, the two recognition sequences for *E. coli* RNA polymerase (12) can be found: TTGCTT 18 base pairs TACCAT (Fig. 4). The strength of this promoter sequence has been determined by the method of Mulligan et al. (15). A score of 46.7% was found, which is relatively low.

Expression of DHDPS. As already reported (22), increases in the *dapA* gene copy number lead to increases in DHDPS activity. As with the chromosomal gene, expression from multicopy plasmids is not subject to regulation in response to variation of the lysine level in the growth medium (data not shown).

Table 2 gives the value of enzyme-specific activities obtained with strain RDA8 (22) harboring various plasmids carrying the *dapA* gene. The same 1.2-kb *Bst*NI fragment carrying *dapA* is present either in pUC9 in both orientations

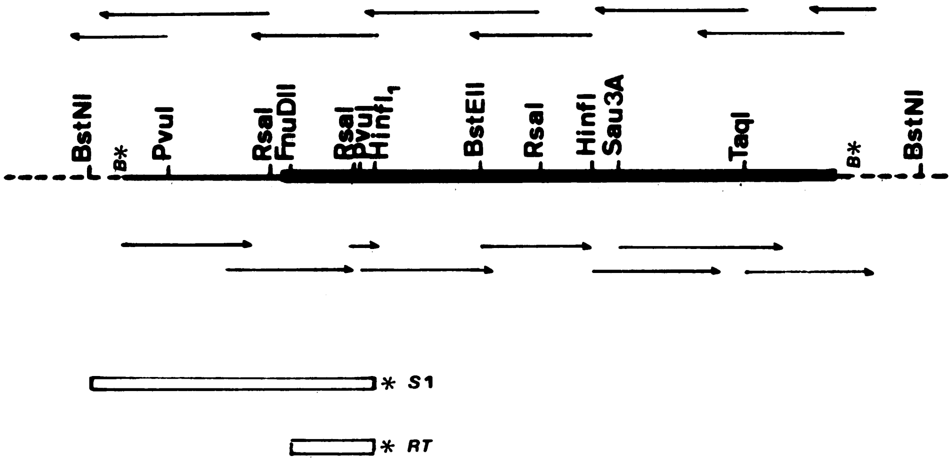

FIG. 3. Sequencing strategy for the *dapA* gene. Only the restriction sites used for the 5′ end labeling of plasmid pDA2 are shown. The arrows indicate the direction and extent of sequence analysis. The heavy bar indicates the *dapA* coding sequence; the dashed bar indicates pUC9 DNA. *B** indicates the position of the *Bst*NI sites lost from pDA2 during cloning in the *Hin*cII site of pUC9. For the determination of the *dapA* mRNA start, probes used were 5′ end labeled at *Hin*fI$_1$. The second restriction cut was with *Fnu*dII for the reverse transcriptase extension method (**RT*) and with *Bst*NI for the S1 mapping method (**S1*).

or in pBR322 in both orientations. DHDPS is expressed in pDA2 and pDA3 at the same level, whatever the orientation of the *dapA* gene with regard to the *lac* promoter (Table 2). This level is twice that obtained with the pBR322 hybrid plasmid pDA1 (22), pDA4, and pDA5 (Table 2), in agreement with the higher copy number of pUC9 compared with pBR322. The DHDPS activity was inhibited 85 to 90% by lysine in vitro in preparations from strains with either

```
                                         50
CCAGGCGACTGTCTTCAATATTACAGCCGCAACTACTGACATGACGGGTGATGGTGTTCACAATTCCACGGCGATCGGCACCCAACGCAGT

100                                               150                                               200
GATCACCAGATAATGTGTTGCGATGACAGTGTCAAACTGGTTATTCCTTTAAGGGGTGAGTTGTTCTTAAGGAAAGCATAAAAAAAACATGCATACAACAATCAGAACGGTTCTGTCTGC

                                    250                                               300
TTGCTTTTAATGCCATACCAAACGTACCATTGAGACACTTGTTTGCACAGAGGATGGCCCATGTTCACGGGAAGTATTGTCGCGATTGTTACTCCGATGGATGAAAAAGGTAATGTCTGT
                                                                MetPheThrGlySerIleValAlaIleValThrProMetAspGluLysGlyAsnValCys
          350                                               400                                               450
CGGGCTAGCTTGAAAAAACTGATTGATTATCATGTCGCCAGCGGTACTTCGGCGATCGTTTCTGTTGGCACCACTGGCGAGTCCGCTACCTTAAATCATGACGAACATGCTGATGTGGTG
ArgAlaSerLeuLysLysLeuIleAspTyrHisValAlaSerGlyThrSerAlaIleValSerValGlyThrThrGlyGluSerAlaThrLeuAsnHisAspGluHisAlaAspValVal
                                                  500                                               550
ATGATGACGCTGGATCTGGCTGATGGGCGCATTCCGGTAATTGCCGGGACCGGCGCTAACGCTACTGCGGAAGCCATTAGCCTGACGCAGCGCTTCAATGACAGTGGTATCGTCGGCTGC
MetMetThrLeuAspLeuAlaAspGlyArgIleProValIleAlaGlyThrGlyAlaAsnAlaThrAlaGluAlaIleSerLeuThrGlnArgPheAsnAspSerGlyIleValGlyCys
                    600                                               650
CTGACGGTAACCCCTTACTACAATCGTCCGTCGCAAGAAGGTTTGTATCAGCATTTCAAAGCCATCGCTGAGCATACTGACCTGCCGCAAATTCTGTATAATGTGCCGTCCCGTACTGGC
LeuThrValThrProTyrTyrAsnArgProSerGlnGluGlyLeuTyrGlnHisPheLysAlaIleAlaGluHisThrAspLeuProGlnIleLeuTyrAsnValProSerArgThrGly
700                                               750                                               800
TGCGATCTGCTCCCGGAAACGGTGGGCCGTCTGGCGAAAGTAAAAAATATTATCGGAATCAAAGAGGCAACAGGGAACTTAACGCGTGTAAACCAGATCAAAGAGCTGGTTTCAGATGAT
CysAspLeuLeuProGluThrValGlyArgLeuAlaLysValLysAsnIleIleGlyIleLysGluAlaThrGlyAsnLeuThrArgValAsnGlnIleLysGluLeuValSerAspAsp
                              850                                               900
TTTGTTCTGCTGAGCGGCGATGATGCGAGCGCGCTGGACTTCATGCAATTGGGCGGTCATGGGGTTATTTCCGTTACGACTAACGTCGCAGCGCGTGATATGGCCCAGATGTGCAAACTG
PheValLeuLeuSerGlyAspAspAlaSerAlaLeuAspPheMetGlnLeuGlyGlyHisGlyValIleSerValThrThrAsnValAlaAlaArgAspMetAlaGlnMetCysLysLeu
          950                                               1000                                              1050
GCAGCAGAAGAACATTTTGCCGAGGCACGCGTTATTAATCAGCGTCTGATGCCATTACACAACAAACTATTTGTCGAACCCAATCCAATCCCGGTGAAATGGGCATGTAAGGAACTGGGT
AlaAlaGluGluHisPheAlaGluAlaArgValIleAsnGlnArgLeuMetProLeuHisAsnLysLeuPheValGluProAsnProIleProValLysTrpAlaCysLysGluLeuGly
                                                  1100                                              1150
CTTGTGGCGACCGATACGCTGCGCCTGCCAATGACACCAATCACCGACAGTGGTCGTGAGACGGTCAGAGCGGCGCTTAAGCATGCCGGTTTGCTGTAAAGTTTAGGGAGATTTGATGGC
LeuValAlaThrAspThrLeuArgLeuProMetThrProIleThrAspSerGlyArgGluThrValArgAlaAlaLeuLysHisAlaGlyLeuLeuOch                 MetAl
```

FIG. 4. Sequence of the *dapA* region. The nucleotide sequence is that of the noncoding strand. The deduced amino acid sequence for the correct reading frame is shown below. Numbering starts at the first letter of the *Bst*NI site (*B** in Fig. 3). The putative ribosome-binding sites are underlined. The wavy arrow indicates the beginning of mRNA; the corresponding promoter signals are underlined twice.

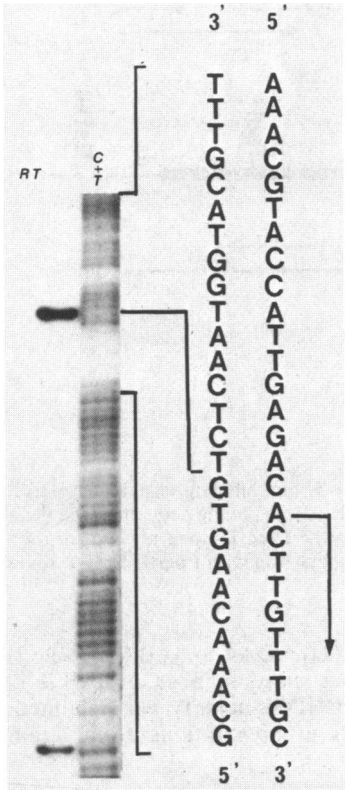

FIG. 5. Identification of the *dapA* transcription start. The primer (118 base pairs) used for the reverse transcriptase extension method is shown in Fig. 3 (RT). The extended fragment is 163 nucleotides long; it is shown along with the sequencing reaction products of the *Hin*fI$_1$-*Bst*NI fragment in Fig. 3.

single-copy chromosomal expression or multicopy plasmids carrying the 1.2-kb insert; this strongly suggests that this 1.2-kb fragment encodes a protein with all the properties (activity and feedback inhibition) expected of DHDPS.

ACKNOWLEDGMENTS

We thank B. Bachmann for taking care of *E. coli* strains, J. Van Eijenhordt, C. Parsot, P. Stragier, and J. Bouvier for helpful discussions, and R. Haselkorn for critical reading of the manuscript.

This work was supported by grants from the Centre National de la Recherche Scientifique (LA136).

LITERATURE CITED

1. **Bachmann, B. J.** 1983. Linkage map of *Escherichia coli* K-12, edition 7. Microbiol. Rev. **47**:180–230.
2. **Berk, A. J., and P. A. Sharp.** 1977. Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1 endonuclease digested hybrids. Cell **12**:721–732.
3. **Bolivar, F., R. L. Rodriguez, P. J. Greene, M. C. Betlach, H. L. Heineker, and H. W. Boyer.** 1977. Construction and characterization of new cloning vehicles. II. A multipurpose cloning system. Gene **2**:95–113.
4. **Boy, E., and J.-C. Patte.** 1972. Multivalent repression of aspartic semialdehyde dehydrogenase in *Escherichia coli* K-12. J. Bacteriol. **112**:84–92.
5. **Boy, E., C. Richaud, and J. C. Patte.** 1979. Multiple regulation of DAP decarboxylase synthesis in *E. coli* K12. FEMS Microbiol. Lett. **5**:287–290.
6. **Bukhari, A. I., and A. L. Taylor.** 1971. Genetic analysis of diaminopimelic acid- and lysine-requiring mutants of *Escherichia coli*. J. Bacteriol. **105**:844–854.
7. **Butour, J. L., B. Felenbok, and J. C. Patte.** 1974. Synthesis of dihydrodipicolinate synthetase in *Escherichia coli* K12. Ann. Microbiol. (Paris) **125**:459–462.
8. **Cameron, J. R., S. M. Panasenko, I. R. Lehman, and R. W. Davis.** 1975. *In vitro* construction of bacteriophage λ carrying segments of the *Escherichia coli* chromosome: selection of hybrids containing the gene for DNA ligase. Proc. Natl. Acad. Sci. USA **72**:3416–3420.
9. **Clarke, L., and J. Carbon.** 1972. A colony bank containing synthetic colE1 hybrid plasmids representative of the entire *E. coli* genome. Cell **9**:91–99.
10. **Cohen, G. N., J. C. Patte, P. Truffa-Bachi, C. Sawas, and M. Doudoroff.** 1965. Repression and end product inhibition in a branched pathway, p. 243–253. *In* Mécanismes de régulation des activités cellulaires chez les microrganismes. Centre National de la Recherche Scientifique, Paris.
11. **Daucé Le Reverend, B., M. Boitel, A. M. Deschamps, J. M. Lebeault, K. Sano, K. Takinami, and J. C. Patte.** 1982. Improvement of *E. coli* strains overproducing lysine using recombinant DNA techniques. Eur. J. Appl. Microbiol. Biotechnol. **15**: 227–231.
12. **Hawley, D. K., and W. R. McClure.** 1983. Compilation and analysis of *E. coli* promoter DNA sequences. Nucleic Acids Res. **11**:2237–2255.
13. **Losick, R., and C. Gilvarg.** 1966. Effect of α-acetylation on utilization of lysine oligopeptides in *E. coli*. J. Biol. Chem. **240**:2340–2346.
14. **Maxam, A., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
15. **Mulligan, M., D. Hawley, R. Entriken, and W. McClure.** 1984. *Escherichia coli* promoter sequences predict *in vitro* RNA polymerase selectivity. Nucleic Acids Res. **12**:789–800.
16. **Neidhardt, F. C., V. Vaughn, T. A. Phillips, and P. L. Bloch.** 1983. Gene-protein index of *Escherichia coli* K-12. Microbiol. Rev. **47**:231–284.
17. **Novick, R. P.** 1962. Micro-iodometric assay for penicillinase. Biochem. J. **83**:236–240.
18. **Parker, J.** 1984. Identification of the *purC* gene product of *Escherichia coli*. J. Bacteriol. **157**:712–717.
19. **Patte, J. C.** 1983. Diaminopimelate and lysine, p. 213–228. *In* K. Hermann and R. Somerville (ed.), Amino acid biosynthesis and genetic regulation. Addison-Wesley Publishing Co., Inc., Reading, Mass.
20. **Patte, J. C., T. Loviny, and G. N. Cohen.** 1962. Repression de la décarboxylase de l'acide meso αε-diaminopimelique par la lysine chez *E. coli*. Biochim. Biophys. Acta. **58**:359–360.
21. **Richaud, C., F. Richaud, C. Martin, C. Haziza, and J. C. Patte.** 1984. Regulation of expression and nucleotide sequence of the *E. coli dapD* gene. J. Biol. Chem. **259**:14824–14829.
22. **Richaud, F., C. Richaud, C. Haziza, and J. C. Patte.** 1981. Isolement et purification de genes d'*E. coli* K12 impliqués dans la biosynthese de la lysine. C. R. Acad. Sci. Paris Ser. III **293**:507–512.
23. **Shedlarski, J. C., and C. Gilvarg.** 1970. The pyruvate aspartic semialdehyde condensing enzyme of *E. coli*. J. Biol. Chem. **245**:1362–1373.
24. **Shine, J., and L. Dalgarno.** 1974. The 3′ terminal sequence of *E. coli* 16s ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. Proc. Natl. Acad. Sci. USA **71**:1342–1346.
25. **Sollner-Webb, B., and R. H. Reeder.** 1979. The nucleotide sequence of the initiation and termination sites for ribosomal RNA transcription in *X. laevis*. Cell **18**:485–499.
26. **Stadtman, E. R., G. N. Cohen, G. Lebras, and H. de Robichon-Szulmajster.** 1961. Feedback inhibition and repression of aspartokinase activity in *E. coli* and *Saccharomyces cerevisiae*. J. Biol. Chem. **236**:2033–2038.
27. **Vieira, J., and J. Messing.** 1982. The pUC plasmids, an M13mp7 derived system for insertion mutagenesis and sequencing universal primers. Gene **19**:259–268.
28. **Yugari, Y., and C. Gilvarg.** 1962. The condensation step in diaminopimelate synthesis. J. Biol. Chem. **240**:4710–4716.