Regulation of Expression and Nucleotide Sequence of the *Escherichia coli* dapD Gene*

Catherine Richaud‡, François Richaud, Christine Martin§, Catherine Haziza, and Jean-Claude Patte§

From the Institut de Microbiologie, Bât 409, Université Paris-Sud, 91405 Orsay Cedex, France

(Received for publication, June 4, 1984)

Regulation of the *Escherichia coli* dapD gene involved in diaminopimelate and lysine biosynthesis was unknown as no convenient enzymatic assay was available until recently. This gene was cloned into pBR322 from a λ transducing phage; its complete nucleotide sequence was established. This sequence shows that from an unknown as no convenient enzymatic assay was available, tetrahydrodipicolinate N-succinyltransferase which catalyzes the third step of the specific lysine-diaminopimelate pathway. The transcriptional start of the *dupD* sequence was established. This sequence shows that from a locus (6), and this region seems to be quite complex; first of all, one mutant of this locus (6) has a different phenotype (slow growth on L broth-rich medium even in the absence of diaminopimelate); secondly, from a previous report of the cloned *dupD* gene (11), more than one cistron could be involved (12).

We report here the identification of the *dupD* gene product and the precise localization of this gene on the chromosomal fragment that we isolated (10) from a λgt11 bacteriophage library constructed by Thomas et al. (13). We also present the nucleotide sequence of the *dupD* gene and its transcriptional start. The regulation of expression of the *dupD* gene was studied using in vitro operon fusion.

**EXPERIMENTAL PROCEDURES**

**Media, Strains, and Plasmids**—Bacterial strains and plasmids are listed in Table I. Bacteria were grown either in L broth-rich medium or in 63 minimal medium supplemented with the required metabolites and 0.4% glucose as the carbon source (18). For plasmid-harboring bacteria, appropriate antibiotics were added (ampicillin, 50 μg/ml; tetracycline, 10 μg/ml). For enzymatic studies, repression conditions were obtained by addition of 4 mM lysine during growth. Conditions leading to lysine limitation were obtained in *dapB* leaky mutant RDB16 (10) grown in minimal medium. In these conditions, doubling time was 4 h (while in diaminopimelate and lysine excess, doubling time was 1 h). Bacteria were harvested after 16 h of growth.

**DNA Manipulations**—Transformation of *E. coli* was done according to Davis et al. (19) after treatment of cells with calcium chloride. Large-scale purification of plasmid DNA was made according to Huennefey et al. (20) after chloramphenicol amplification (21). For DNA sequencing, plasmids were further purified on sucrose gradient (22). For rapid analysis of recombinant plasmids, the alkaline lysis procedure of Birnboim and Doly (23) was used. Restriction enzymes (New England Biolabs and Boehringer Mannheim) were used according to Ref. 19. Lignations were performed with T4 DNA ligase (Boehringer Mannheim) in 20 mM Tris-HCl, pH 7.5, 10 mM MgCl₂, 10 mM dithiothreitol, and 1 mM EDTA, and 100 mM NaCl. For rapid analysis of recombinant plasmids, the alkaline lysis procedure of Birnboim and Doly (22) was used.

Restriction enzymes (New England Biolabs and Boehringer Mannheim) were used according to Ref. 19. Lignations were performed with T4 DNA ligase (Boehringer Mannheim) in 20 mM Tris-HCl, pH 7.5, 10 mM MgCl₂, 10 mM dithiothreitol, and 1 mM EDTA. Incubation was done overnight at 4 °C in 10 μl.

Analytical or preparatory electrophoresis was done in 10 mM Tris-borate, pH 8.3, 2.5 mM EDTA on agarose or polyacrylamide gels.

All DNA sequencing was done using the Maxam and Gilbert method (23). Purified plasmid was cut with appropriate restriction enzymes and 5' end-labeled using the reaction reaction of Berkner and Folk (24). After secondary restriction cleavage, the labeled fragments were purified on polyacrylamide gel electrophoresis electrophoresed by the method of McDaniel et al. (25). Products of chemical degradation reactions were analyzed with the denaturing gels of Sanger and Coulson (26).

Mapping of the 5' end of transcripts was done by two methods: protection against SI nuclease (Röhringer Mannheim) degradation (27) and primer extension with reverse transcriptase (gift from M. Yamiv, Institut Pasteur, Paris) (28).

The 9 genes involved in diaminopimelate and lysine biosynthesis in *Escherichia coli* are of particular interest because they are scattered on the chromosome. The expression of 5 genes (lysC (1), asd (2), dupB (3), dupE (4), and lysA (5)) is known to be regulated by the intracellular lysine pool (4). There is no evidence that regulation by lysine should be dependent on a common regulatory element. These 5 genes are currently studied in our laboratory to compare their regulatory sequences.

For two other genes of the regulon, two genetic loci were identified by Bukhari and Taylor (6) in the 4-min region of the chromosome and were arbitrarily called *dupD* and *dupD*. These genes should encode, respectively, tetrahydrodipicolinate and succinylaminopimelate transferase (7). These enzymes of the lysine-diaminopimelate pathway were previously characterized by Gilvarc and colleagues (8, 9); however, substrates for these enzymes were not readily available, and the exact functional roles of the *dupC* and *dupD* loci were not identified.

In the case of the so-called *dupD* locus several mutations are known (6, 10), and this region seems to be quite complex; first of all, one mutant of this locus (6) has a different phenotype (slow growth on L broth-rich medium even in the absence of diaminopimelate); secondly, from a previous report of the cloned *dupD* gene (11), more than one cistron could be involved (12).

The 9 genes involved in diaminopimelate and lysine biosynthesis, 6 are known to be lysine regulated.

* This work was supported by the Centre National de la Recherche Scientifique (LA 136) and by a grant from Roussel-UCLAF (for C.H.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ To whom correspondence should be addressed.

§ Present address, Laboratoire de Chimie Bactérienne, Centre National de la Recherche Scientifique, BP 71, 13277 Marseille Cedex 9, France.
was done from appropriate strains: either from chromosomal dapD gene of strain RM4102 (14) or from the same strain harboring multicopy plasmid pDD2 carrying the dapD gene. Hybridization was done during 3.5 h at 35°C in the presence of 80% formamide. Enzymatic treatments were then carried out as described by Sollner-Webb and Reeder (28).

**Enzyme Assays**—Enzymatic determinations were made from cell-free crude extracts. Galactokinase assays were performed according to McKenney et al. (16); β-lactamase specific activities were measured as described in Chenais et al. (29).

**RESULTS**

**Precise Localization and Identification of the dapD Gene**—Precise localization of the dapD gene was performed on a 5.5-kilobase pair EcoRI bacterial fragment carrying this gene. This fragment was subcloned in both orientations from a λ bacteriophage into plasmid pBR322 to give plasmids pDD1 and pDD3 (10). Its restriction map is in agreement with a previously reported map of this region (12) except for the rightmost EcoRI site of our chromosomal fragment which is absent in ApolU9-transducing phage (12); other results from our laboratory indicate that EcoRI sites on the borders of the insert could be either created (dapD, this work; dapB (30)) or lost (dapA1) probably due to EcoRI* conditions used during the λgAB library construction.

Construction of plasmids pDD2, pDD4, and pDD5 is shown in Fig. 1. These plasmids complement all dapD mutant strains: dapD4 and dapD12 from Bukhari and Taylor (6), and RDD3, RDD7, RDD15, RDD22, and RDD32 from our laboratory (10). These results show that the dapD gene defined by these 7 mutations is localized in the 1300-bp AluI fragment of pDD5. Enzymatic studies (31) were performed on one of these strains, RDD32, which was found to lack any tetrahydridopicolinate N-succinylase activity. Conversely, crude extracts from this strain harboring plasmid pDD5 were highly enriched in tetrahydridopicolinate N-succinylase activity as compared to a wild-type strain.

**Nucleotide Sequence of the dapD Gene**—The nucleotide sequence of the 1.3-kilobase pair AluI fragment was determined using the method of Maxam and Gilbert (23); the strategy is shown in Fig. 2. The complete sequence on both strands from HindIII site on the left part of the fragment (bp 1) to the last BstNI site on the right end (bp 1177) is given in Fig. 3. A large open reading frame of 274 triplets is found from ATG (bp 217) to ochre codon TAA (bp 1039); this open reading frame is preceded by a GAG sequence 7 nucleotides upstream from the initiating ATG, which could be part of a ribosome-binding site by complementarity of 16 S RNA (32).

Codon usage has been compared to that of total tryptophan operon coding sequences (33). For most of the amino acids the most abundant tRNA in the cells is used, for instance, CUC for Leu, ACC for Thr, CGU or CGA for Arg, GAA for Glu. Exceptions can be noticed for Gln and Tyr where CAG or CUG is favored over CAA and UAU. The molecular weight of 30,040 for the polypeptide deduced from DNA sequencing is in good agreement with the 31,000 value described for the subunit of tetrahydridopicolinate N-succinyltransferase (31) and the presence of a 33-kDa poly-
peptide in *E. coli* maxicells harboring plasmid pDD2 (data not shown).

Immediately downstream of the TAA nonsense codon a typical ρ-independent terminator sequence (34) is found in the 3' flanking region, a GC-rich inverted repeat followed by a run of 7 thymidylate residues in the nontemplate DNA strand. 33 bp upstream from the dapD gene initiating ATG (bp 217) a TGA codon (bp 184) closes a reading frame opened for at least 60 triplets (limit of the sequence so far determined) that could be the end of an adjacent gene.

**Determination of the Transcriptional Start**—The 5' end of the dapD messenger RNA has been localized by nuclease S1 mapping assay (24) and by the reverse transcriptase extension method (25). The strategy followed is shown in Fig. 2. As may be seen in Fig. 4, both methods localize the transcriptional start at nucleotides G and A (bp 187 and 185). Thus, a leader sequence of only 29-31 nucleotides is found before the translational initiation ATG (bp 217).

**Upstream of the transcriptional start** the perfect '-35' RNA polymerase recognition site TTGACA (34) is found at the correct distance. As for the other important recognition sequence, the 'Pribnow box' (ideally TATAAT), two weak signals are found either AACGAT (with 3 matches out of 6),
Escherichia coli dapD Gene

The nucleotide sequence shows one open reading frame of 822 bp beginning by ATG (preceded by GAG) and terminated by TAA. Transcription start has been identified 30 ± 1 bp before the translational start, and a very strong putative signal sequence of the galK gene. After transformation of the galK strain N100 (16), bacteria harboring this plasmidic fusion pCJ5 were selected as red clones on MacConkey galactose agar plates in the presence of ampicillin.

Table II

| Strain          | Minimal mediuma | Excess lysineb (4 mM) |
|-----------------|-----------------|----------------------|
| pCJ5/N100       | 101             | 52                   |
| pCJ5/RDB16      | 358c            | 64                   |

*Specific activities are expressed as nanomoles of galactose 1-phosphate produced/min/mg of protein in the crude extract, normalized to the plasmid copy number; plasmid copy number was calculated from β-lactamase specific activities as described in Chenais et al. (29).

bGrowth conditions.

cIn this dap leaky strain, minimal medium leads to lysine limitation with a doubling time of 4 h. Lysine limitation is controlled by derepression of lysine-sensitive aspartokinase; these conditions are only 30% derepressive as compared to the maximal level of aspartokinase found in chemostat.

**DISCUSSION**

We report here the nucleotide sequence of the E. coli dapD gene located in the 4-min region of the chromosome (6, 7).

**FIG. 5. In vitro construction of an operon fusion between the dapD and galK genes.** Heavy line is bacterial insert. The Rsal fragment (Rsal1-Rsal2 from pDD2) was cloned in the unique HincII site of pUC5 plasmid (15) coding for a-domain of the lacZ gene. This intermediate plasmid, pCJ3, was selected as Lac+ clones in lacZ donor strain JM83 (15). Then the HindIII-BamHI fragment from pCJ3 (containing the 706-bp RsaI fragment) was cloned between the HindIII and BamHI sites of plasmid pK04 (16) containing the coding sequence of the galK gene. After transformation of the galK strain N100 (16), bacteria harboring this plasmidic fusion pCJ5 were selected as red clones on MacConkey galactose agar plates in the presence of ampicillin.

**FIG. 4. Identification of the start point for dapD gene transcription.** The protected fragments after treatment with SI nuclease or reverse transcriptase (RT) are shown along with the sequencing reaction products of HindII-HincII fragment (see Fig. 2).

15 bp from the "-35" sequence and 11 nucleotides from the mRNA start, or GATAAA (with 4 matches out of 6), 18 bp from the "-35" sequence and 8 nucleotides from the mRNA start.

**Regulation of Expression of the dapD Gene—** In order to study eventual regulation of the expression of dapD gene, in vitro fusion with the galK gene was performed. Plasmid pK04, carrying a promoterless galK gene, has been designed by McKenney et al. (16) to analyze DNA fragments promoting procaryotic transcription. Such fragments can be introduced into unique sites, which places the galK-coding region under the control of the promoter present in the inserted fragment. It is assumed that the amount of galactokinase produced is linearly related to the transcriptional efficiency of the inserted promoter (16).

Restriction sites upstream from and inside the dapD gene did not allow direct construction of an operon fusion with the galK gene. So a two-step procedure was used (see Fig. 5). The constructed fusion, pCJ5, contains a 706-bp RsaI fragment beginning 370 bp before translational initiation ATG and 305 bp before the "-35" transcriptional recognition sequence.

Promoter expression and regulation were further characterized by quantitative measurements of galactokinase synthesis in vitro. Enzyme levels were determined in two strains and in different growth conditions ranging from lysine excess to partial limitation. Results presented in Table II show that expression of the dapD gene as measured with thedapD-galK plasmidic fusion pCJ5 is subject to lysine regulation. A 5-fold difference could reproducibly be obtained between lysine excess and lysine limitation.
mutant is different from all other dupD mutants. Our experiments (data not shown) localized this mutation in the 5' region of the gene. Slow growth on LB-rich medium in the absence of diaminopimelate may account for the leakiness of this mutant. Though phenotypically different from all other dupD mutants, it is located in the dupD gene which actually consists of only one cistron. This gene encodes the tetrabody- dipicolinate N-succinyltransferase which catalyzes the third step of the specific lysine-diaminopimelate pathway and not the succinyl-diaminopimelate aminotransferase as was previously arbitrarily admitted (6, 7).

Results obtained with in vitro fusion between the dupD promoter and the galK structural gene allow the conclusion that expression of the dupD gene is regulated by lysine. The 5-fold range regulation observed between lysine excess and lysine limitation (only partial in our experiments) may reflect a larger phenomenon; we cannot exclude the possibility that a regulatory element present in limiting amount is involved. Experiments are in progress to obtain single-copy chromosomal fusions for further study of dupD expression.

Thus, out of the 7 genes of the lysine pathway studied so far in our laboratory (4), only the expression of dapA is not subject to repression by lysine. As for the other genes of the same regulon, examination of the sequence of the dupD gene does not reveal any of the characteristics of an attenuation-like structure (36). The sequence immediately upstream from the transcriptional start may be compared to the camesus promoter which is poorly expressed. Some of the data presented here do not fulfill this prediction: (i) galactokinase expression from plasmid pCJ5 carrying a hybrid dupD-gulK operon is relatively high compared with data from "ideal" promoters constructed by De Boer et al. (40) or with the asd gene promoter which is known to be well expressed; (ii) when mapping the dupD gene transcriptional start, protected hybrids could be readily detected even with mRNA extracted from bacteria harboring only one copy of the dupD gene per genome (see Fig. 4). To correlate this high in vivo expression of the dupD gene with its weak promoter sequence, we propose an activation mechanism like the one demonstrated for the lysA gene (31). Activation of the dupD gene promoter could be modulated by the internal lysine pool. Experiments are currently in progress in our laboratory to answer this question.

Acknowledgments—We are indebted to C. Gilvarg for his generous help in identification of the dupD gene product. We wish to thank J. Bouvier, P. Stragier, and A. Kerr for helpful discussions and critical reading of the manuscript and F. Mareclet for his kind help with computer analysis of the sequence.

REFERENCES

1. Stadtman, E. R., Cohen, G. N., Le Bras, G., and de Robichon-Szulmajster, H. (1961) J. Biol. Chem. 236, 2033-2038

2. Cohen, G. N., Patte, J. C., and Boezi, J. (1963) C. R. Hebd. Acad. Sci. Paris 256, 1398-1392

3. Temari, H., and Gilvarg, C. (1974) J. Biol. Chem. 249, 3034-3040

4. Patte, J. C. (1983) in Amino Acid Biosynthesis and Genetic Regulation (Herrmann, K., and Somerville, R., eds) pp. 213-228, Addison-Wesley, Reading, MA

5. Patte, J. C., Loviny, T., and Cohen, G. N. (1965) Biochim. Biophys. Acta 99, 523-530

6. Bokhari, A. I., and Taylor, A. L. (1971) J. Bacteriol. 105, 844-845

7. Bachmann, B. (1983) Microbiol. Rev. 47, 180-230

8. Gilvarg, C. (1961) J. Biol. Chem. 236, 1429-1431

9. Peterkovsky, B., and Gilvarg, C. (1961) J. Biol. Chem. 236, 1432-1435

10. Richaud, F., Richaud, C., Haziza, C., and Patte, J. C. (1981) C. R. Acad. Sci. Paris Ser. II 293, 507-512

11. Friesen, J. D., Parker, J., Watson, R. J., Bendiai, D. S., Reeh, S. V., Pedersen, S., and Fili, N. P. (1976) Mol. Gen. Genet. 148, 93-98

12. Bendiai, D. S., and Friesen, J. D. (1981) Mol. Gen. Genet. 181, 356-362

13. Thomas, M., Cameron, J. R., and Davis, R. W. (1974) Proc. Natl. Acad. Sci. U. S. A. 71, 4579-4583

14. Patte, J. C., Morand, P., Boy, E., Richaud, C., and Borne, F. (1980) Mol. Gen. Genet. 179, 319-325

15. Vieira, J., and Messing, J. (1982) Gene 19, 259-268

16. McKenney, K., Shimatake, H., Court, D., Schneissner, U., Brady, C., and Rosenberg, M. (1981) in Gene Amplification and Analysis, (Chirkijian, J., and Papas, T., eds) Vol. II, pp. 383-415, Elsevier/North-Holland, New York

17. Bolivar, F., Rodriguez, R. L., Greene, P. J., Betlach, M. C., Heyneker, H. L., and Boyer, H. W. (1977) Gene 2, 95-113

18. Miller, J. H. (1972) in Experiments in Molecular Genetics, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York

19. Davis, R. W., Botstein, D., and Roth, J. R. (1980) in Advanced Bacterial Genetics, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York

20. Humphreys, G. O., Willshaw, G. A., and Anderson, E. S. (1975) Biochim. Biophys. Acta 383, 457-463

21. Clewell, D. B. (1972) J. Bacteriol. 110, 667-676

22. Birnboim, H. C., and Doly, J. (1979) Nucleic Acids Res. 7, 1515-1523

23. Maxam, A. M., and Gilbert, W. (1980) Methods Enzymol. 85, 495-560

24. Berkner, K. L., and Folk, W. R. (1977) J. Biol. Chem. 252, 3176-3184

25. McDonell, M. W., Simon, M. N., and Studier, F. W. (1977) J. Mol. Biol. 110, 119-146

26. Sanger, F., and Coulson, A. R. (1978) FERS Lett. 170, 107-110

27. Park, A. J., and Sharp, P. A. (1977) Cell 12, 721-722

28. Sollner-Webb, B., and Reeder, R. H. (1979) Cell 18, 485-499

29. Chenas, J., Richaud, C., Ronaeray, J., Surdin-Kerjan, Y., Cherest, H., and Patte, J. C. (1981) Mol. Gen. Genet. 182, 456-461

30. Bouvier, J., Richaud, C., Richaud, F., and Stragier, P. (1984) J. Biol. Chem. 259, 2734-2741

31. Simms, S. A., Voige, W. H., and Gilvarg, C. (1984) J. Biol. Chem. 259, 1523-1533

32. Shine, J., and Dalgarno, L. (1974) Proc. Natl. Acad. Sci. U. S. A. 71, 1342-1346

33. Yanofsky, C., Platt, T., Crawford, I. P., Nichols, B. P., Christie, G. E., Horowitz, H., Van Cleemput, M., and Wu, A. M. (1981) Nucleic Acids Res. 9, 6647-6668

34. Rosenberg, M., and Court, D. (1979) Annu. Rev. Genet. 13, 319-355

35. Pribnow, D. (1975) Proc. Natl. Acad. Sci. U. S. A. 72, 784-788

36. Yanofsky, C. (1981) Nature (Lond.) 289, 751-758

37. Hawley, D. K., and McClure, R. M. (1983) Nucleic Acids Res. 11, 2237-2255

38. Stefanou, J. E., and Gralla, J. D. (1982) Proc. Natl. Acad. Sci. U. S. A. 79, 1069-1072

39. Mulligan, M. E., Hawley, D. K., Entrenik, R., and McClure, W. R. (1984) Nucleic Acids Res. 12, 789-800

40. De Boer, H. A., Comstock, L. J., and Vasser, M. (1983) Proc. Natl. Acad. Sci. U. S. A. 80, 21-25

41. Stragier, P., Richaud, F., Borne, F., and Patte, J. C. (1983) J. Mol. Biol. 168, 307-320

*C. Richaud and C. Martin, unpublished results.