Structure and Autoregulation of the metJ Regulatory Gene in Escherichia coli*

Isabelle Saint-Girons†, Nathalie Duchange, Georges N. Cohen, and Mario M. Zakin

From the Unité de Biochimie Cellulaire, Département de Biochimie et Génétique Moléculaire, Institut Pasteur, 28 Rue du Docteur Roux, 75724 Paris Cedex 15, France

The nucleotide sequence of the Escherichia coli metJ regulatory gene (312 nucleotides) has been determined as well as that of two mutations located within the gene. Analysis of the sequence downstream from the metJ gene has revealed inverted repeats homologous to several intercistronic regions, also reported to occur between operons. A hybrid protein that contains the 55 first amino acid residues of the metJ protein substituting for the 8 amino acid residues at the NH₂ terminus of β-galactosidase was produced by gene fusion. The hybrid protein retaining β-galactosidase activity was purified. Its amino-terminal sequence was determined and this allowed us to locate the translational start codon of the metJ gene. Evidence was provided for autoregulation by repression of the metJ gene. By sequencing upstream from metJ, the region situated between the metJ and metB genes was found to contain putative operator structures that we propose to call "Met boxes."

The levels of the proteins involved in methionine biosynthesis are elevated in strains of Escherichia coli bearing mutations in the metK or in the metJ gene (1). It appears that metJ codes for a regulatory protein which, when combined with methionine (or one of its derivatives), causes repression of the expression of the methionine regulon. The seven methionine structural genes are scattered on the E. coli chromosome (2) and are independent units of transcription, except in the case of two genes (metB and metL) which are arranged in an operon (3, 36). The metJ aporepressor could thus interact with several operator loci.

Here, we present the complete nucleotide sequence of the metJ regulatory gene as well as that of two mutations located within the gene. We address also the question of autoregulation by repression of the metJ gene. To determine whether the product of metJ regulates its own synthesis, we have constructed in vitro a metJ-lacZ hybrid gene which leads to a hybrid protein endowed with β-galactosidase activity. In this system, the synthesis of β-galactosidase is used as an index of transcription from the metJ promoter. We have been able to investigate the quantitative effect of metJ alleles on expression of β-galactosidase activity in such metJ+ derivatives of the strains bearing a single copy of the metJ-lacZ hybrid gene.

MATERIALS AND METHODS

Construction of a metJ-lacZ Hybrid Gene—We made use of the pMC1403 plasmid suitable to detect fragments with transcription and translation start signals (4). A 754-bp AluI-AluI fragment of pMAD4 (positions 589 to +165 of Fig. 1), spanning the metJ promoter (3) and 165 bp of the metJ gene, was inserted into the Smal site of pMC1403. The MC1000 strain (ΔmetE) transformed by the recombinant plasmid obtained (pIP24) gave rise to colonies expressing β-galactosidase activity. This indicates that a hybrid gene has been constructed coding for a chimeric protein with the NH₂ terminus of the metJ repressor and the COOH terminus of β-galactosidase. The expression of the hybrid protein is under the control of the metJ promoter on the pIP24 multicopy plasmid. The structure of the hybrid gene was verified by restriction enzyme analysis and DNA-sequencing experiments.

The hybrid gene was transferred to a λ bacteriophage. An EcoRI-SacI II fragment of the recombinant plasmid carrying the hybrid gene was inserted between the left arm of λSEG (5) and the right arm of λGT4 (6). The recombinant bacteriophages were selected by detection of β-galactosidase activity in the plaques. These plaques were purified three times and used to lysogenize a λαc-argE strain (X1a100C su). The lysogen was made metJ by cotransduction of the metJ185am marker (carried by J1190) with argE. The metJ genotype was recognized by its norleucine resistance and methionine excretion (7). The lysogenic strain was then transformed with the pMAD4 multicopy plasmid to produce a strain with several copies of the wild-type metJ gene.

Purification of the Hybrid Protein, Product of the metJ-lacZ Hybrid Gene—The MC1000 strain transformed by the pIP24 recombinant plasmid was used as a source of the pmetJ-lacZ hybrid protein. The purification is a slight modification of the procedure of Fowler and Zabin (8). Cultures were grown in a 15-liter fermenter in LB broth supplemented with ampicillin (50 μg/ml) and the bacterial pellet was frozen at -30 °C. All subsequent steps were performed in the presence of 1 mM phenylmethylsulfonyl fluoride. 30 g of bacterial cells were broken by ultrasonic treatment. A fourth step involving a Sepharcl S300 chromatography in the presence of 6 M urea was added to the purification scheme described (8). This procedure gave a protein pure enough as judged by polyacrylamide gel electrophoresis in the presence of sodium dodecyl sulfate to be subjected to automated sequencing.

NH₂ Terminal Sequence Determination of the pmetJ-lacZ Hybrid Protein—Approximately 8 nmol of pure pmetJ-lacZ were dialyzed against several liters of 50 mM ammonium bicarbonate. The sample was lyophilized and applied to the spinning cup of a modified SOCOSI-PS100 sequenator. A high-performance liquid chromatography system (Waters) was used to identify the phenylthiohydantoin amino acid derivatives which were eluted with a methanol gradient from a Merck Lichrospher 60 CHS super column.

Enzymes and Chemicals—β-Galactosidase was assayed by hydrolysis of o-nitrophenyl-β-D-galactoside according to Miller (9). DNA polymerase large fragment and restriction endonucleases were obtained from New England Biolabs. T4 polynucleotide kinase was from T4 polynucleotide kinase was from

*This work was supported by the Centre National de la Recherche Scientifique (U. A. 517), the Ministère de l'Industrie et de la Recherche, the Commissariat à l'Énergie Atomique, and the Institut National de la Santé et de la Recherche Médicale (841006). 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 all correspondence should be addressed.

1 The abbreviations used are: bp, base pairs; kb, kilobase pairs; pmetJ-lacZ, translational product of the metJ-lacZ hybrid gene.
Boehringer. Acrylamide was from British Drug Houses, Ltd., urea and boric acid were from Serva, hydrazine was from Eastman, and dimethyl sulfate was from Aldrich. All other chemicals, mostly from Merck, were analytical grade or purer. Phenol was distilled.

Nucleotide Sequence Determination—The nucleotide sequences were determined by the chemical method of Maxam and Gilbert (10).

Bacterial Strains and Growth Media—Media and genetic manipulations were done as described in Miller (9). E. coli K12 strains used are the following: XA100Car is F- argE (amber), Δ(lac, pro), nalaA8, rip8, su-. MC1000 is araD139, Δ(aral, leu)7697, Δlac X74, galU, galK, strA (11). Jl119 is Hfr relA1, spoT, metJam185 (12). PD107 is Hfr metBI, relA1, recBC (pRG1355). PD109 is Hfr metBI, relA1, recBC (pRG1317). The three last strains were a gift of Dr. Ronald Greene, Veterans Administration Hospital, Durham, NC. GT729 is X1100 Car su- (xmetJ-lacZ) (this paper). GT745 is G732 metJam185 (constructed of an arginine-independent transductant).

RESULTS AND DISCUSSION

DNA Sequence of the metJ Gene—The E. coli metJ gene was cloned together with the three other met genes of the metJBLF cluster in plasmid vectors (3, 13). In the pMD4 plasmid, metJ was localized in a DNA segment starting at 5.1 kb and ending at 5.7 kb upstream from a single EcoRI site, which has been shown to be near the right (clockwise) end of the bacterial DNA insert containing the met genes (3, 13). Moreover, our previous structural evidence (3) indicates that metJ must lie within a 0.7-kb DNA segment situated upstream from the origin of the metB structural gene.

The complete nucleotide sequence of the metJ gene is presented (Fig. 1) as well as the deduced corresponding protein sequence. The DNA fragments and the restriction sites used in the sequence determination are indicated in Fig. 2. Analysis of the sequence (Fig. 1) shows that an open reading frame begins at position −141 and ends at position 312. Position −141 is localized precisely at 138 bp (counterclockwise) from the origin of the metB gene. As previously published (3), the metJ gene is transcribed in the opposite direction to that of the metBL and metF transcriptional units.

Two ATG codons in phase (positions −141 and −3) could theoretically be the translational start codon of the metJ structural gene. In order to determine which ATG is used as initiator codon, we determined the N-terminal sequence of the metJ protein. Actually, we constructed a metJ-lacZ hybrid gene (see "Materials and Methods"), purified the corresponding hybrid protein and determined its N-terminal sequence. The results given by automated sequencing are Ala-Glu-Trp-Ser-Gly and allow the unambiguous choice of the ATG in position −3 as initiator codon of metJ. The gene is thus 312 nucleotides long and encodes a single polypeptide chain of 104 residues with a molecular weight of 11,996. This is in agreement with the analysis of the peptides synthesized under the direction of plasmid genes by the maxicell procedure; the metJ product was thus identified in sodium dodecyl sulfate-polyacrylamide gel as a 12,000-dalton radioactive polypeptide (38).

The results of the analysis of the nucleotide sequence of the metJ gene are consistent with the sequence given in Fig. 1. The codon usage in the metJ gene is not strictly characteristic of the codon usage found in weakly expressed genes by Grosjean and Fiers (15). Neither an AT-rich sequence followed by a sequence related to CAATCAA, reported (16) to be characteristic of rho-dependent terminators, nor a structure resembling a rho-independent terminator (16) have been found downstream of metJ. However, at positions 346 to 370 and 411 to 440 (Fig. 1), two potential stem and loop structures can be formed. The nucleotide sequence of these structures is very similar to the consensus structure found in several intercistronic regions (17). Moreover, this DNA palindromic segment has been recently reported to occur also between operons (18), and it has been postulated that it could be involved in the regulation of transcription termination events. Further experiments are needed to confirm if these DNA segments are indeed responsible for the transcription termination of the metJ gene.

Analysis of the Deducing Protein Sequence and Comparison with Other Regulatory Proteins—Analysis of the protein sequence indicates that, although the total number of basic and acidic residues are of the same order, their distribution is very different. 67% of the basic residues are present in the N-terminal half of the protein and 68% of the acidic ones are in the C-terminal half. Also, 8 out of the 15 basic residues are

Fig. 1. Sequence of the metJ gene and its flanking regions. The given sequence is that of the noncoding strand. The deduced amino acid sequence for the correct reading frame is shown. The putative operator sequence presented the usual dyad symmetry is boxed (see text). SD, ribosomal attachment site. • and †, positions of the met Jam185 and met J184 mutations, respectively. Nucleotide 1 corresponds to the first amino acid of the isolated protein. The exact position of each number is indicated by the nucleotide concerned. The arrow indicates the two palindromic domains found in the 3′ region of the gene. For clarity purposes, metJ has been shown here in the opposite direction to the counterclockwise direction it actually occupies on the E. coli chromosome.

**Structure and Autoregulation of E. coli metJ Regulatory Gene**
Structure and Autoregulation of E. coli metJ Regulatory Gene

Fig. 2. Sequencing strategy and restriction map of the part of the pBR322-met (pMAD4) hybrid plasmid (13) corresponding to the metJ gene. The arrows indicate the sites used for labeling as well as the direction and extent of the sequence. More than 90% of the sequence has been established on both strands. The following abbreviations have been used for restriction enzymes: B, Bacillus; S, Sau3A1; M, MspI; B, BstNI; D, Ddel; N, NcoI; H, Hinfl. The orientation of metJ is the same as that of Fig. 1.

**Table 1**

| Strain       | Source of metJ* | Number of metJ* copy | β-Galactosidase specific activity |
|--------------|-----------------|----------------------|----------------------------------|
| GT745        | 0               | 0                    | 1                                |
| GT732        | Chromosome      | 1                    | 0.42                             |
| GT732 (pMAD4) | Chromosome + plasmid | ≤24               | 0.15                             |

*The specific activity for GT745 given in minimal medium was 632 units of β-galactosidase (defined as nanomoles of o-nitrophenol produced per min and per mg of protein). This activity was arbitrarily chosen as 1. The same activities were found in minimal medium supplemented with 5 mM methionine.

present in tandem mainly in the N-terminal region of the polypeptide sequence.

Comparison of the amino acid sequences of several prokaryotic regulatory proteins has revealed two discrete regions of homology, region 1 and region 2 (19). The results obtained when wild-type and mutant proteins were compared suggest that the homologous sequences of these two regions could play an important role in the interaction with DNA. Moreover, both regions of homology are located in the known DNA-binding domains of lacI (for a review, see Refs. 20 and 21), cl (22), and crp (23, 24) products. In the so-called region 1 of homology (19), the most homologous segment is the T-V(or I)-S(or G)-R-tetrapeptide totally conserved in crp, lacI, galR, arac, and lycR products and to a lesser extent in trpR, lexA, and trpR proteins. When the amino acid sequence of the metJ product is compared with that of the regulatory protein sequences mentioned, a tripeptide T-V-S characteristic of region 1 is found at positions 25, 26, and 27, corresponding to the nucleotides 73 to 81 of Fig. 1. The position of the tripeptide in the N-terminal part of the metJ protein, together with the predominance of basic residues in the same part may indicate that the N-terminal region of the metJ protein interacts with the DNA target sequence for the met repressor.

DNA Sequence of Two metJ Mutations—The metJ184 and metJ185 alleles responsible for a derepressed synthesis of the metJ proteins were isolated, in strain J100, by selection for ethionine resistance (12). The pRCG135 and pRCG137 recombinant plasmids were constructed (38) by cloning in pBR322, using the Xmet102 metJ184 and Xmet102 metJ185 transducing bacteriophages as sources of mutant alleles of metJ, respectively (12, 25). In addition, it was previously shown that the metJ185 defect must be due to an amber mutation (12).

Nucleotide sequence of the structurally altered metJ genes indicates that, in the pRCG137 DNA, the G base at position 8 (Fig. 1) is changed into an A, transforming the Trp codon into an amber stop codon; in the pRCG135 DNA insert, the only modification is situated at position 178 (G → A; Fig. 1), transforming the Ala codon into a Thr codon. These results are consistent with the pattern obtained by analysis of the peptides synthesized under the direction of plasmid genes by the maxicell procedure (38). Indeed, the radioactive polypeptide identified as the metJ product is present in a reduced amount in extracts of maxicells containing the pRCG135 plasmid and absent in those with the pRCG137 plasmid (metJ185). Identification of the structural change and absence of gene product confirm the genetic evidence that the metJ185 allele is an amber mutation.

The reduced amount of the metJ product obtained in the pRCG135 maxicell preparation may indicate that the mutation (Als → Thr) has modified the stability of the metJ polypeptide, perhaps rendering it more accessible to proteolytic degradation.

Table 1 gives data for a metJ-lacZ fusion strain, in which the lacZ gene fused to the metJ promoter, providing strong evidence for autoregulation. The data also suggest that the repression of the metJ protein synthesis is insensitive to supplementation of the medium with concentrations of methionine that normally repress the synthesis of the methionine structural genes (26).

It seems that autoregulation of regulatory genes is found in a great number of cases. Synthesis of the araC product has been shown to be self-regulated (27), as have the lexA gene (28), the λI gene (29), and the hutC gene of Salmonella typhimurium (30). However, the maltT product is not auto-regulated (31). Self-repression of the synthesis of repressors of biosynthetic pathways was shown for the first time in the case of the trpR gene (32, 37) and an operator-like sequence in the 5’ part of the trpR gene has in fact been identified (32, 33).

In our case, analysis of the 5’ part of the metJ gene shows a regulatory region for two divergent transcriptional units (the metJ gene and the metBL operon). The promoter of the metBL operon has been identified and the transcription start is located at position −147 of Fig. 1 (3). The results presented by Duchange et al. (3) show the presence of another promoter activity in an orientation opposite to the metBL operon, but do not allow to locate precisely the metJ promoter. This must await the determination of the transcription start of metJ. A plausible operator region (3) was proposed for the metBL operon (position −174 to −195 of Fig. 1) which was homologous to a DNA segment located in the 5’ flanking regions of
FIG. 3. Organization of the metJBLF cluster of E. coli K12. The DNA fragment is 5735 bp. The size of the genes and of the intercistronic regions are indicated in base pairs. The name of the enzymes encoded by the genes and the molecular weight of their subunit are in the lower part of the scheme. In the upper part of the scheme, the direction of transcription of the genes and the distance between the transcription start signal and the start codon of each transcriptional unit are indicated (the question mark indicates that the distance is unknown). RNAm, messenger ribonucleic acid. K, kilodaltons. In this figure, metJ is oriented from right to left (counterclockwise in the E. coli chromosome), in opposition to Figs. 1 and 2, where only the metJ segment is analyzed.

the metF gene (14) and of the metA gene. Since metJ is transcribed on the other DNA strand and in the opposite direction to metB, the putative operator region of the metBL operon could also be that of the metJ gene. An example of a gene cluster constituted of two divergent operons with an internal common operator region has been already described (34). By analogy with the arg region where the homologous operator regions were called "Arg boxes," we propose "Met boxes" although no operator-constitutive mutations have been reported yet where these sequences would be altered. It is possible that regulation at the level of transcription is not the only regulation affecting metJ expression.

CONCLUSION

With the information presented in this paper and previously, we are able to give the complete organization of the metJBLF cluster of E. coli. The nucleotide sequences of the metB, metF, and metJ genes are known by previous reports of this laboratory (3, 14, 35) and that of metJ is presented here. As indicated in Fig. 3, the three independent transcriptional units are organized as follows: the metJ structural gene, transcribed in the E. coli chromosome (2) counterclockwise occupies 312 bp and the metBL and metF structural genes (clockwise) occupy 3585 and 888 bp, respectively. A complex regulatory region is found between the metJ and metB structural genes. The metJBLF cluster represents 5735 nucleotides from the last base of the second palindromic unit of metJ (position 440 of Fig. 1) to the last base of the putative rho-dependent terminator of metF (14). The transcription start sites for the metBL operon and for the metF gene were determined. Finally, the results presented here indicate that the metJ regulatory gene is autoregulated.

Acknowledgments—We would like to thank Drs. B. Blazy and J. Hozmayer for helpful discussions concerning protein purification and Drs. Gicquel-Sanzey and R. Greene for profitable discussions and advice. We thank Drs. M. Springer and R. Greene for the gift of bacteriophage DNA and bacterial strains, respectively. We thank Drs. A. Ullmann and N. Guiso-Maclof for preliminary experiments of purification of the hybrid protein. I. Park has kindly performed some DNA-sequence experiments. The expert technical assistance of D. Margarita is acknowledged. We are indebted to Drs. M. Bruschi and M. Bonicel for rapid N-terminal sequence determination. We thank L. Girardot for typing the manuscript.

REFERENCES

1. Flewin, M. (1976) in Metabolic Pathways (Greenberg, D. M., ed) 3rd Ed., Vol. 7, pp. 457-505, Academic Press, New York.
2. Bachmann, B. J. (1980) Microbiol. Rev. 44, 190-230.
3. Drouhet, N., Zakin, M. M., Ferrara, P., Saint-Girons, I., Park, I., Tran, V. S., Py, M. C., and Cohen, G. N. (1983) J. Bacteriol. 258, 14504-14509.
4. Galfre, M. J., Chou, J., and Cohen, S. N. (1980) J. Bacteriol. 143, 971-980.
5. Enquist, L., Tiemeier, D., Leder, P., Weisberg, R., and Sternberg, N. (1976) Nature (Lond.) 260, 356-359.
6. Davis, R. W., Botstein, D., and Roth, J. R. (1980) Advanced Bacterial Genetics, p. 237, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
7. Cohen, G. N., and Jacob, F. (1980) C. R. Acad. Sci. Paris 248, 1490-1492.
8. Johnson, J. R., and Zakin, M. M. (1983) J. Bacteriol. 157, 413-419.
9. Miller, J. H. (1972) in Experiments in Molecular Genetics, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
10. Maxam, A. M., and Gilbert, W. (1977) Proc. Natl. Acad. Sci. U. S. A. 74, 560-564.
11. Cassadan, M. J., and Cohen, S. N. (1980) J. Mol. Biol. 138, 179-207.
12. Liljestrand-Golden, C. A., and Johnson, J. R. (1984) J. Bacteriol. 157, 413-419.
13. Zakin, M. M., Greene, R. C., Ducret-Yavetz, A. Duchange, N., Ferrara, P., Py, M. C., Margarita, D., and Cohen, G. N. (1983) Mol. Gen. Genet. 187, 101-106.
14. Saint-Girons, I., Duchange, N., Zakin, M. M., Park, I., Margarita, D., Ferrara, P., and Cohen, G. N. (1985) Nucleic Acids Res. 11, 6723-6732.
15. Groisman, E., and Fiers, W. (1982) Gene 18, 199-202.
16. Rosenberg, M., and Court, D. (1979) Annu. Rev. Genet. 13, 319-353.
17. Higgins, C. F., Ferro-Luzzi Ames, G. F., Barnes, W. M., Clement, J. M., and Hofnung, M. (1982) Nature (Lond.) 290, 760-762.
18. Gilson, E., Clement, J., and Frutting, D., and Hofnung, M. (1984) EMBO J. 3, 1417-1421.
19.Gilson, E., Clement, J., and Hofnung, M. (1983) Mol. Gen. Genet. 187, 101-106.
20. Beleyreuther, K. (1978) in The Operon (Miller, J. H., and Remnikoff, W., ed) pp. 125-154, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
21. Miller, J. (1978) in The Operon (Miller, J. H., and Remnikoff, W., eds) pp. 31-38, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
22. Page, C. O. (1969) Nature (Lond.) 222, 858-860.
23. Kelley, L. Girardot for typing the manuscript.
24. Debarbouille, M., and Schwartz, M. (1979) J. Mol. Biol. 132, 521-534.
25. Singleton, C. R., Roeder, W. D., Bogosian, G., Somerville, R. L., and Weith, H. L. (1980) Nucleic Acids Res. 8, 1551-1560.
26. Gunashe, R. P., and Yanofsky, C. (1980) Proc. Natl. Acad. Sci. U. S. A. 77, 7117-7121.
27. Piette, J., Curin, R., Boyen, A., Chaffet, D., Crabbe, M., Van Vliet, F., Glansdorff, N., Squires, C., and Squires, C. L. (1982) Nucleic Acids Res. 10, 3001-3004.
28. Zakin, M. M., Duchange, N., Ferrara, P., and Cohen, G. N. (1983) J. Bacteriol. 258, 3029-3035.
29. Greene, R. C., and Smith, A. A. (1984) J. Bacteriol. 159, 767-769.
30. Kelley, L. R., and Yanofsky, C. (1982) Proc. Natl. Acad. Sci. U. S. A. 79, 3120-3124.
31. Smith, A. A., and Greene, R. C. (1984) J. Bacteriol. 259, 14279-14281.

2 E. Ron, personal communication.