Nucleotide Sequence of the purM Gene Encoding 5′-Phosphoribosyl-5-aminimidazole Synthetase of Escherichia coli K12*

(Received for publication, November 18, 1985)

John M. Smith and Henry A. Daum III
From the Department of Biochemistry and Molecular Biology, Louisiana State University School of Medicine, Shreveport, Louisiana 71130

5′-Phosphoribosyl-5-aminimidazole synthetase (EC 6.3.3.1), encoded by the purM gene of Escherichia coli, catalyzes the synthesis of 5′-phosphoribosyl-5-aminimidazole from 5′-phosphoribosylformylglycinamidine. The purM gene was subcloned from the Clarke and Carbon (Clarke, L., and Carbon, J. (1976) Cell 9, 91-99) plasmid pLC1-41 and the nucleotide sequence determined. The mature protein, as deduced from the purM structural gene sequence, contains 344 amino acid residues and has a calculated Mr of 36,726. The 5′ end of the purM mRNA was determined by mung bean nuclease mapping to be 44-45 nucleotides upstream of the proposed GTG translation initiation codon. A C-G-rich region characteristic of stringently controlled promoters is located immediately in front of the proposed purM promoter region. Comparison of the upstream sequences of the purM and the coregulated purF loci revealed a highly conserved (33 of 39 base pairs are identical) sequence. The presumptive purM promoter is located in this region, thus suggesting that both purine loci share a common mechanism of regulation mediated through this sequence.

5′-Phosphoribosyl-5-aminimidazole synthetase (EC 6.3.3.1) catalyzes the ATP-dependent synthesis of 5′-phosphoribosyl-5-aminimidazole (AIR) from 5′-phosphoribosylformylglycinamidine (FGAM) (1) as shown by the following equation.

\[ \text{FGAM} + \text{ATP} \rightarrow \text{AIR} + \text{ADP} + \text{P}_i \]

This reaction catalyzes the fifth step, i.e. the imidazole ring closure, in the de novo synthesis of the purine nucleotides. 5′-Phosphoribosyl-5-aminimidazole synthetase has been partially purified (10-fold) and characterized from pigeon liver (2). Recently, 5′-phosphoribosyl-5-aminimidazole synthetase from chicken liver has been purified to homogeneity and shown to be part of a multifunctional protein (M, ~110,000) which also contains glycaminide ribonucleotide synthetase and glycaminide ribonucleotide transformylase, the second and third enzymes in the pathway, respectively (41). This is not the case in Escherichia coli since the purification and subsequent characterization of 5′-phosphoribosyl-5-aminimidazole synthetase shows only the single 5′-phosphoribosyl-5-aminimidazole synthetase activity (3). 5′-Phosphoribosyl-5-aminimidazole synthetase is encoded by the purM locus in E. coli (4) while in Salmonella typhimurium, it is the product of the analogous purF locus (5).

In E. coli and S. typhimurium, the purM and purF genes along with the genes for the other steps in the purine biosynthetic pathway have been shown to be coregulated (29). The purine biosynthetic pathway and the purF locus in particular has also been shown to be stringently regulated (30). Additionally, a DNA-binding protein has been isolated and shown to exhibit GTP-dependent binding to plasmids carrying the purF and purM genes while exhibiting ATP-dependent binding to plasmids containing the purF and purA genes (42). However, the precise mechanism of this regulation remains to be elucidated.

In this paper we report the cloning and DNA sequence of a 2.0-kb DNA fragment containing the E. coli purM structural gene. The mature protein, as deduced from the purM structural gene sequence, contains 344 amino acid residues and has a calculated Mr of 36,726. The 5′ end of the purM mRNA was determined by mung bean nuclease mapping to be 44-45 nucleotides upstream of the proposed GTG translation initiation codon. A C-G-rich region associated with stringently controlled promoters (37) is located immediately in front of the proposed purM promoter region. Comparison of the upstream sequences of the purM and the coregulated purF loci revealed a highly conserved (33 of 39 bp are identical) sequence. The presumptive promoters of both purM and purF are located in this region, thus suggesting that the common mechanism of coregulation is mediated through this sequence.

**EXPERIMENTAL PROCEDURES**

**Strains and Media**—E. coli K12 strain TX939 (ara Δ (lac) purM srlC300::Tn10 recA56) was constructed and used as a recipient for the isolation of purM plasmids by complementation. Strain TX939 containing plasmid pLS18 was used as a source of purM mRNA. Strain JM83 (6) was used as a recipient to identify subclones containing restriction fragments from the purM region while strain JM101 (6) was employed for the propagation of M13 bacteriophages. Strains were made competent and transformed by the procedure of Dargent and Ehrlich (7). The minimal medium of Neishardt et al. (8) and the rich media described by Miller (9) were used for the growth of the E. coli K12 strains.

**Plasmids**—Plasmid pLC1-41 from the Clarke and Carbon library (10) was obtained from the Coli Genetic Stock Center, Yale University, New Haven, CT. Plasmids pUC9 (12), pUC12 (13), and pUC15 (13) were employed for the subcloning of restriction fragments from plasmid pLC1-41. Plasmid pNM481 (28), a promoter-translation detection plasmid, was employed to create an in-phase purM-lacZ fusion. Restriction fragment sizes were determined by the program of Schaffer (14) that was translated into Pascal and run on an Apple II Plus computer.

*This work was supported by National Institutes of Health Grant AI 20068. Computer resources used to carry out our studies were provided in part by the BIONET National Computer Resource for Molecular Biology funded by National Institutes of Health Grant U14 RR-01685-02. 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.

*The abbreviations used are: kb, kilobase pair(s); bp, base pair(s).
Nucleotide Sequence of purM

Plasmid | PurM Complementation
--- | ---
PjLC1-41 | +
PjJS16 | +
PjJS18 | +
PjJS24 | +
PjJS79 | -

Fig. 1. Partial restriction map and complementation analysis of plasmid pLC1-41. The thick line represents chromosomal DNA, and the thin line represents ColEl DNA. Only the restriction sites pertinent to the purM locus and this study are shown. Complementation studies were carried out in strain TX393.

![Restriction Map of Plasmid pLC1-41](image)

FIG. 2. Restriction endonuclease sites and sequencing strategy for purM. The location of the major 6-base pair restriction enzyme sites in purM are shown at the top of the figure. The arrows denote sequencing direction, and the length of the arrow is proportional to the number of nucleotides determined. The PstI site is designated as the 5' end. The purM-coding region extends from nucleotide 780 to nucleotide 1814 and is indicated by the thickened line.

DNA Isolation—The procedures previously described (15) were employed.

Cloning of purM—Plasmid pLC1-41 from the Clarke and Carbon library (10) has been previously identified as containing the purM locus (11, 42). A preliminary restriction map of pLC1-41 was determined for PstI, EcoRI, and SmaI restriction enzymes (Fig. 1). The total size of pLC1-41 was determined to be approximately 18 kb which reflects a chromosomal insert size of 11.6 kb. The plasmid pLC1-41 was tested and found not to carry either of its flanking loci, guaA or upp. DNA fragments generated by the PstI restriction enzyme were subcloned into pUC9 and transformed into strain TX393 (purM). All purM+ plasmids examined from purM+ transformants contained a common 10.5-kb PstI fragment which was found in both orientations. By comparison with our derived restriction map, it was concluded that this 10.5-kb PstI fragment additionally contained ColEl DNA and spanned the fusion point between the bacterial DNA and the original ColEl vector. A plasmid, pJS16, containing the 10.5-kb PstI fragment was then subjected to further restriction analysis in order to more precisely locate the purM structural gene. A 1.8-kb SphI fragment failed to complement strain TX393, but a 1.8-kb HincII fragment did complement. This localized the purM gene to the 1.8-kb HincII restriction fragment. The results are summarized in Fig. 1.

DNA Sequence Analysis—DNA sequences were determined by the dideoxy chain termination method of Sanger et al. (16). A 539-bp PstI-HincII restriction fragment containing upstream DNA was also subjected to a similar procedure. The resulting fragments were then ligated into the appropriate cloning sites in M13mp8 and M13mp9 (17) or M13mp10 and M13mp11 (13) and transformed into JM101 (13). Colorless plaques were individually picked and propagated for the preparation of DNA (18). The DNA sequences were compiled and analyzed by computer (19) and in part by the facilities of the BIONET resource.

RNA Studies—RNA was extracted from strain TX393 containing pJS18 by the sodium dodecyl sulfate/hot phenol method (20). The S1 nuclease procedure of Berk and Sharp (21) was used to determine the 5' end of the purM mRNA except that mung bean nuclease (34, 36) instead of S1 nuclease was employed to degrade single-stranded nucleic acids. A 317-bp HincII-Hinfl fragment that spanned the end of the purM locus was 5' end labeled with [γ-32P]ATP and polynucleotide kinase after the fragment had been dephosphorylated with calf intestinal alkaline phosphatase (22). mRNA from strain TX393 was hybridized to the labeled fragment for 8 h at 49 °C. Sixty units of mung bean nuclelease were then added and the mixture incubated for 15 min at 37 °C. After termination of the reaction by phenol extraction, the mung bean nuclease digestion mix was ethanol precipitated, resuspended in loading buffer, and loaded onto a sequencing gel. The size of the protected fragment was determined after autoradiography by comparison to an accompanying Sanger dideoxy sequencing ladder. The fragment size was corrected by 0.5 nucleotide for the phosphate group (23).
Enzymes and Chemicals—[^22]PdATP was obtained from New England Nuclear, and [y-^32]P]ATP was obtained from ICN (Irvine, CA). T4 DNA ligase, DNA polymerase I (Klenow fragment), T4 polynucleotide kinase, calf intestinal alkaline phosphatase, and SmaI and Asp718 restriction enzymes were obtained from Boehringer Mannheim. All other enzymes were obtained from New England Biolabs (Beverly, MA). Mung bean nuclease, deoxy- and dideoxynucleotide triphosphates were obtained from Pharmacia P-L Biochemicals. 5-Bromo-4-chloro-3-indoyl-β-D-galactoside and all other chemicals were obtained from Sigma.

RESULTS AND DISCUSSION

Nucleotide Sequence of the DNA Fragment Encoding purM—The entire sequence of a 2.043-kb DNA fragment that contained the purM gene was determined for both strands from overlapping DNA fragments. A detailed restriction map and the specific DNA fragments sequenced are shown in Fig. 2. The DNA sequence shown is from the PstI site to a TaqI site distal to the end of the purM coding region and is numbered from the first nucleotide of the PstI site as the 5' end (Fig. 3). The sequence shown includes 779 nucleotides of upstream sequence as well as 229 nucleotides of sequence distal to the proposed purM coding region. A potential Shine-Dalgarno sequence (26) for the purM coding region that obeys rules 1 and 2 of Stormo et al. (27) is located at nucleotide position 769–771 (Fig. 3). The upstream sequence contains an open reading frame from nucleotide 1 to 566 of unknown function or potential expressibility. However, the algorithm of Fickett (24) predicts that it is a coding sequence. Analysis of this partial open reading frame for rare codon usage (39) indicates that with the exception of the rare codon for threonine (ACA), rare codon usage is that of a poorly expressed gene. While it is not followed by a typical rho-independent terminator sequence (25), the coding region does end in a region of interrupted inverted symmetry (554–597).

Derived Amino Acid Sequence—The DNA sequence shown in Fig. 3 contains an open reading frame of 1,050 nucleotides from nucleotide 765 to nucleotide 1814. This open reading frame is initiated with a GTG codon while another in phase GTG codon is located downstream at nucleotides 780–782. The first 20 residues of the mature 5'-phosphoribosyl-5-aminoimidazole synthetase, identified by automated Edman degradation, correspond to nucleotides 783–842 with threonine as the NH₂-terminal residue (3). The GTG codon at nucleotides 780–782 is preceded by an acceptably spaced (5–9 bases) Shine-Dalgarno sequence (GGA) according to rules 1 and 2 of Stormo et al. (27) while the GTG codon at nucleotides 765–767 is not. Thus, based on the presence of a Shine-Dalgarno sequence we propose that the GTG codon at nucleotides 780–782 is the translational initiation codon and that the corresponding initiation residue is processed to yield threonine as the NH₂-terminal residue of the mature peptide. However, other mechanisms involving alternate translational

[^22]: The DNA sequence data reported here have been submitted to GenBank™.
initiations and subsequent processing have not been rigorously ruled out. Therefore, based on the identified amino acid residues and as deduced from the DNA sequence, the mature 5′-phosphoribosyl-5-aminimidazole synthetase contains 344 amino acid residues and has a calculated $M_r$ of 36,726. This value is in good agreement with the reported molecular weight of purified 5′-phosphoribosyl-5-aminimidazole synthetase of 33,000 (3). If the KpnI site at nucleotide position 981–986 is cleaved with the isoschizomer Asp718 and the 5′ overhang filled in, it is predicted to form an in-phase lacZ fusion with the promoter-translation detection plasmid pNM481 (28). Such a plasmid was constructed and found to produce an active β-galactosidase. The expected structure of the purM-lacZ fusion point was confirmed by DNA sequencing after subcloning into M13. Based on these criteria, we believe that the DNA sequence of the purM locus is correct. Analysis of the purM coding sequence for rare codon usage (39) indicated the codon usage pattern is typical of a highly expressed gene. Additionally, an effort was made to identify the ATP-binding site by comparison to the proposed ATP-binding domain of adenylate kinase. The first 63 amino acid residues that contain the ATP-binding domain of E. coli adenylate kinase (40) were compared to the amino acid sequence of 5′-phosphoribosyl-5-aminimidazole synthetase by the ALIGN program of BIONET set to the default values. No extensive areas of homology were found.

**RNA Flanking Regions: Mung Bean Nuclease Studies**—The 5′ end of the purM mRNA was determined by mung bean nuclease mapping. A 317-bp HincII-HinfI fragment that spanned the end of the purM gene was 5′ end labeled with polynucleotide kinase and [γ-32P]ATP after dephosphorylation with calf intestinal alkaline phosphatase according to the procedures of Maniatis et al. (22). RNA extracted from strain TX393 containing plasmid pJS18 was hybridized to the labeled fragment. After mung bean nuclease treatment, the protected fragment was sized on a DNA-sequencing gel using a Sanger dye sequencing ladder as a standard. A protected fragment of 119 or 120 nucleotides was detected on the sequencing gel (Fig. 4). Based on this size and the strong resemblance to the purF control region (see below), the probable transcription initiation nucleotide is 736 or 737 if RNA polymerase spacing constraints prevail (Fig. 3).

18 bp beyond the end of the purM structural gene is a region of dyad symmetry that has the potential to form a hairpin loop if translated into RNA (Fig. 5). This potential structure has a calculated free energy of −10.2 kcal/mol and might act as a rho-independent terminator (25) for purM transcription.

**purM Control Region**—Based on the mRNA-mapping studies, transcription initiation occurs 44 or 45 bp upstream of the start of the purM structural gene (Fig. 3). Within this region the sequence TAGAATT at position 1833–1836 shows the main fragment protected (Lane 3) shows the main fragment protected RNA from strain TX393 containing plasmid pJS18 was hybridized to the protected fragment. After mung bean nuclease treatment, the protected fragment was sized on a DNA-sequencing gel using a Sanger dye sequencing ladder as a standard. A protected fragment of 119 or 120 nucleotides was detected on the sequencing gel (Fig. 4). Based on this size and the strong resemblance to the purF control region (see below), the probable transcription initiation nucleotide is 736 or 737 if RNA polymerase spacing constraints prevail (Fig. 3).

18 bp beyond the end of the purM structural gene is a region of dyad symmetry that has the potential to form a hairpin loop if translated into RNA (Fig. 5). This potential structure has a calculated free energy of −10.2 kcal/mol and might act as a rho-independent terminator (25) for purM transcription.

**purM Control Region**—Based on the mRNA-mapping studies, transcription initiation occurs 44 or 45 bp upstream of the start of the purM structural gene (Fig. 3). Within this region the sequence TAGAATT at position 1833–1836 shows the main fragment protected (Lane 3) shows the main fragment protected RNA from strain TX393 containing plasmid pJS18 was hybridized to the protected fragment. After mung bean nuclease treatment, the protected fragment was sized on a DNA-sequencing gel using a Sanger dye sequencing ladder as a standard. A protected fragment of 119 or 120 nucleotides was detected on the sequencing gel (Fig. 4). Based on this size and the strong resemblance to the purF control region (see below), the probable transcription initiation nucleotide is 736 or 737 if RNA polymerase spacing constraints prevail (Fig. 3).

18 bp beyond the end of the purM structural gene is a region of dyad symmetry that has the potential to form a hairpin loop if translated into RNA (Fig. 5). This potential structure has a calculated free energy of −10.2 kcal/mol and might act as a rho-independent terminator (25) for purM transcription.

**purM Control Region**—Based on the mRNA-mapping studies, transcription initiation occurs 44 or 45 bp upstream of the start of the purM structural gene (Fig. 3). Within this region the sequence TAGAATT at position 1833–1836 shows the main fragment protected (Lane 3) shows the main fragment protected RNA from strain TX393 containing plasmid pJS18 was hybridized to the protected fragment. After mung bean nuclease treatment, the protected fragment was sized on a DNA-sequencing gel using a Sanger dye sequencing ladder as a standard. A protected fragment of 119 or 120 nucleotides was detected on the sequencing gel (Fig. 4). Based on this size and the strong resemblance to the purF control region (see below), the probable transcription initiation nucleotide is 736 or 737 if RNA polymerase spacing constraints prevail (Fig. 3).

18 bp beyond the end of the purM structural gene is a region of dyad symmetry that has the potential to form a hairpin loop if translated into RNA (Fig. 5). This potential structure has a calculated free energy of −10.2 kcal/mol and might act as a rho-independent terminator (25) for purM transcription.

A C-G-rich sequence (GCGCCG) at position 731–736 is characteristic of promoters under stringent control (37). This is in agreement with the reports that the purine biosynthetic pathway and the purF operon is under stringent control (30). However, the purM gene has not been specifically tested for its stringent response.

**Identification of a Common Sequence in the purM and purF Control Regions**—Since the purM gene has been shown to be coregulated with the purF, purE, and purJHD loci (29), the purM upstream sequence was compared with the upstream sequence of the purF operon (30). A common sequence that was highly conserved (35 of 39 bp) was found to be present in the upstream regions of both purine loci (Fig. 6 (30)). The differences between the two conserved sequences consists of two transitions, three transversions, and a one base deletion.
Both promoters are located within this sequence, and this suggests that the common regulation of the purM and purF loci is achieved through this sequence. A DNA-binding protein with specificity for plasmids carrying the purF and purM genes has been described (42), and this common sequence may represent the site of binding. These two regions are aligned for comparison in Fig. 6, and the transcription initiation points as well as the translational initiation codons are also indicated. Additionally, the transcription initiation sites and the conserved sequences for both loci are located approximately the same distance from the initial translational start points. Further studies will be necessary to genetically confirm the identity of the -10 regions of both loci and to determine the mechanism of their regulation.

The purM control region was also compared to the guaBA operon control region (31, 38), and no obvious points of similarity were found. This is not unexpected since the guaBA operon control region is regulated differently from the de novo purine biosynthetic pathway (29, 32). However, if the other de novo purine loci share the same regulatory mechanism, then this conserved sequence should be found in their control regions. Other genes of the purine biosynthetic pathway are being currently sequenced to determine the structure of their control regions.

In summary, the purM and purF control regions exhibit two main points of similarity. (i) In both genes the distance from the transcription initiation point to the coding region is approximately 40 bp. (ii) A highly conserved DNA sequence (33 of 39 bp) is present at the same location in both control regions.

Acknowledgments—The skillful typing of this manuscript by Carol Ann Ardoin is appreciated. The excellent technical assistance of M. Susan Edgerton and Laura Packer is also acknowledged.

REFERENCES