THE JOURNAL OF BIOLOGICAL CHEMISTRY
© 1985 by The American Society of Biological Chemists, Inc.

Vol. 260, No. 18, Issue of August 25, pp. 10378–10387, 1985
Printed in U.S.A.

Regulation of *Escherichia coli purF*

ANALYSIS OF THE CONTROL REGION OF A *pur* REGULON GENE*

(Received for publication, January 7, 1985)

Christopher A. Makaroff‡ and Howard Zalkin

From the Department of Biochemistry, Purdue University, West Lafayette, Indiana 47907

Escherichia coli purF **has been determined to be the distal gene of a polycistronic operon. The first gene of the *purF* operon encodes a hydrophobic 17.9-kDa protein of unknown function. Deletion analyses indicate that the 17.9-kDa protein plays no role in the regulation of *purF* in *cis*. mRNA hybridization studies establish that *purF* is regulated at the transcriptional level. Enzyme and mRNA levels are repressed 11–17-fold by excess adenine. A single mRNA start site at nucleotide +1 was identified for transcripts synthesized *in vivo*. Two sites, at +1 and approximately +30, were used for transcription initiation *in vitro*. The *purF* promoter is localized between nucleotides −96 and −7 with sequences upstream of −71 necessary for high level expression. Initial evidence suggests that transcription is subject to stringent control. Deletion analyses localize the *purF* control element to a region between nucleotides −71 and +35. A putative control site between nucleotides −35 to +3 strongly resembles a 5′ flanking sequence in the co-regulated gene *purM*. This site contains an imperfect inverted repeat sequence that is characteristic of sites recognized by regulatory proteins and is a candidate for the *purF* operator. This is the first detailed analysis of a gene involved in *de novo* purine nucleotide biosynthesis.**

The mechanisms for genetic regulation of *de novo* purine nucleotide synthesis in microorganisms are poorly understood. Expression of *pur* regulon genes in *Escherichia coli* and *Salmonella typhimurium* is repressed when purines are added to the medium (1). Analysis of *E. coli* and *S. typhimurium* purine regulatory mutants has suggested that *purE, purF, purM*, and *purJHD* may be controlled by a common regulatory element, assumed to be a *purR*-encoded repressor protein (1). Koduri and Gots reported the isolation of a DNA-binding protein having the specificity expected for a *pur* gene repressor protein (2). However, a direct relationship between *purR* and the *pur*-specific DNA-binding protein has not been established. In fact, the *purR* locus has not, to this date, been mapped.

An alternative approach to study control of *pur* regulon gene expression is to characterize the 5′ control region of representative *pur* genes. With this approach in mind, this laboratory has cloned and sequenced the gene *purF* from *E. coli* (3) and *Bacillus subtilis* (4) and the corresponding gene *ADE4* from *Saccharomyces cerevisiae* (5). In this report we present a detailed analysis of the 5′ control region of *E. coli purF*. The *purF* promoter was localized by identification of the *in vivo* and *in vitro* mRNA start sites. Unexpectedly, the 5′ end of the *purF* transcript was found to encode a protein of 162 amino acids of unknown function, indicating that *purF* is the second cistron in a polycistronic mRNA. Correlation of mRNA and enzyme levels establishes that regulation is at the level of transcription. A deletion analysis has localized the *purF* regulatory element within a 107-bp[1] region overlapping the promoter.

EXPERIMENTAL PROCEDURES

Strains and Media—*E. coli* strains TX158 (3), N100 (6), JM83 (7), and JM103 (7) have been described. RNA was isolated from the wild type strain W3110. Strain CM1001, a *recA* derivative of MC1000 (8) was the host for *purF-lacZ* fusion plasmids. Strain N100 was the host for *purF-galK* fusion plasmids. Strain TX533 (*ara Δlac purR*::Muc*ts*Tn5) was obtained from John M. Smith, Louisiana State University Medical Center.

Minimal growth medium contained salts (9), 0.5% glucose, and 2 μg/ml thiamin. Repression was accomplished by the addition of adenine at 25 μg/ml. L broth was employed as a rich medium. McConkey/galactose (Difco) and minimal/*X*-Gal plates were used to screen cells harboring *purF-galK* and *purF-lacZ* fusion plasmids, respectively.

Plasmid Construction—Table I lists plasmids used in this study and their relevant properties. Four key parental plasmids are diagrammed in Fig. 1. Plasmid pSB2 (3) contains the *purF* structural gene and approximately 2.2 kb of upstream DNA cloned between the *Eco*RI and *Hin*dIII sites of pBR322. Plasmid pCM2 contains a *purF-lacZ* fusion that was constructed by blunt end ligation of a 2.2-kb *Hin*cII fragment from pSB2, Fig. 1, into the *Sma*I site of pMC1403. The *Hin*cII/*Sma*I junction in pCM2 created a translational fusion in which the first 13 codons of *purF* were fused in-frame to *lacZ*. Plasmid pCM11 is a *purF-galK* transcriptional fusion constructed by cloning the 2.2-kb *Eco*RI/*Bam*HI fragment from pCM2 into pK0100. Plasmid pCM13 was constructed by ligating a 311-bp *Taq*I fragment, nucleotides −96 to 216, from pSB2 into the *Acc*I site of pK0100.

Deletions used in this study are diagrammed under the parental plasmids in Fig. 1. A series of deletions of *purF* upstream DNA was constructed in pCM3 and then transferred to pCM2. Deletions in pCM2Δ170 (−27 to 144) and pCM2Δ52 (71 to 123) were obtained by treatment of *Sst*II-linearized pCM3 with S1 nuclease at an enzyme to DNA ratio of 100 units/μg DNA for 30 min at 37 °C. Plasmid pCM2X is a derivative of pCM2 containing *Xba*I linkers ligated into the *Puv*II site at nucleotide 444 (Fig. 2). The linkers were originally inserted into pCM3 and then transferred to pCM2. Plasmids pCM11Δ15 and pCM11Δ16 were constructed by *Bal*31 exonuclease digestion (12) of pCM11 linearized at the *Stu*I site. *Bam*HI linkers were ligated at the deletion end points. The deletion plasmid pCM11Δ12 was obtained by first subcloning the *Eco*RI/*Bam*HI fragment of pCM2X into a pBR322 derivative from which the *Nde*I site

* This work was supported by United States Public Health Service Grant GM24658. This is Journal Paper No. 10,179 from the Purdue University Agricultural Experiment Station. 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.

‡ Recipient of National Institutes of Health Predoctoral Training Grant GM07211.

[1] The abbreviations used are: bp, base pairs; kb, kilobase pairs; kDa, kilodalton; REP, repetitive extragenic palindromic; *X*-Gal, 5-bromo-4-chloro-3-indoyl-β-D-galactoside.

Downloaded from www.jbc.org by on November 9, 2006

TABLE I

Plasmids and phage used to study purF expression

Plasmid	Description
pMC1403	*lacZ* fusion vector (10)
pK01	*galK* fusion vector (6)
pK0100	Polylinker from pUC13 cloned into *Eco*RI/*Hin*dIII sites of pK01
pUC8	Cloning vector (7)
pUC13	Cloning vector (7)
M13mp11	Cloning vector (11)
pSB2	5.3 kb of *E. coli* DNA containing *purF*$^+$ cloned in *Eco*RI and *Hin*dIII sites of pBR322 (3)
pCM2	*purF-lacZ* translational fusion: 2.2 kb of upstream *purF* DNA and first 13 codons of *purF*-coding region cloned into pMC1403
pCM3	*Eco*RI/*Bam*HI fragment from pCM2 cloned into pUC8
pCM2X	*Xba*I linkers ligated at *Pvu*II site of pCM2
pCM2Δ52	pCM2 deletion of nucleotides 71–123 in *purF* DNA
pCM2Δ170	Deletion from pCM3Δ170 transferred to pCM2
pCM3Δ170	Deletion of nucleotides −27–144 in pCM3
pCM11	*Eco*RI/*Bam*HI fragment of pCM2 cloned into pK0100 to yield a *purF-galK* fusion
pCM11Δ12	Deletion of nucleotides 35–444 in pCM11
pCM11Δ15	Deletion of nucleotides −209–−96 in pCM11
pCM11Δ16	Deletion of nucleotides −197–26 in pCM11
pCM13	*Taq*I fragment (nucleotides −96–216) cloned into pK0100
pCM13Δ24	Deletion of nucleotides −96–−72 in pCM13
M13mp11-410	*Nde*I-*Pvu*II fragment (nucleotides 35 to 444) cloned into M13mp11
M13mp11-623	*Stu*I-*Pvu*II fragment (nucleotides −183–44) cloned into M13mp11
M13mp11-822	*Hinc*II fragment (nucleotides 518–1220) cloned into M13mp11

had been removed. The unique *Xba*I and *Nde*I sites were then used to delete intervening DNA. The *Eco*RI/*Bam*HI fragment carrying the deletion was then transferred to pK0100. Plasmid pCM13Δ24 was obtained by linearizing pCM13 with *Eco*RI followed by treatment with exonuclease III and S1 nuclease (13). *Eco*RI linkers were ligated to the deletion end point and used to transfer the deletion to pK0100 on a *Eco*RI/*Bam*HI fragment. All deletion end points were verified by sequencing. All plasmids carry the ampicillin resistance gene, *bla*.

Single stranded DNA probes for mRNA hybridization were obtained by ligating appropriate fragments into the *Hinc*II site of the replicative form of M13mp11. M13 phage were grown in strain JM103 as described (11).

RNA Methods—For nuclease S1 mapping of the 5′ end of the *purF* transcript, RNA was isolated (14) from strain W3110 grown to midlog phase in minimal media. Conditions for S1 nuclease mapping were previously described (15).

RNA for mRNA determinations was isolated (16) from strain TX158 harboring plasmid pSB2 (*purF*$^+$). Cells were grown to midlog phase in minimal media and pulse labeled with [^{3}H]uridine (10 μCi/ml) for 90 s. Repression was achieved by addition of 25 μg/ml adenine. Hybridizations to DNA immobilized on nitrocellulose filters were carried out as described (17). The rate for the first order decay of mRNA was determined as outlined by Hauser and Hatfield (18). At the end of a 90-s pulse, incorporation of [^{3}H]uridine was terminated by the addition of 0.4 mg of rifampicin and 2 mg of unlabeled uridine/ml. Aliquots were withdrawn every 30 s for 2 min and *purF*-specific mRNA quantitated. mRNA levels were corrected for decay (18), and the number of uridine residues in the downstream mRNA segment was normalized to the uridine content of the upstream sequence.

In Vitro Enzyme Synthesis—Coupled transcription/translation was carried out using plasmids pSB2, pCM3, and pCM3Δ170 in the standard S-30 system (19). The S-30 extract was prepared from the *purF* deletion strain TX158. Proteins were radiolabeled with [^{14}C] leucine, precipitated with trichloroacetic acid, and subjected to electrophoresis on sodium dodecyl sulfate-polyacrylamide gels. Radiolabeled proteins were visualized by autoradiography.

DNA Sequence Determination—DNA sequences were determined by the method of Maxam and Gilbert (20) and analyzed by computer (21). The sequence of single-stranded DNA probes was confirmed by dideoxy sequencing (22).

In Vitro Transcription—purF-specific transcripts were synthesized *in vitro* from isolated DNA fragments containing the *purF* promoter. Transcripts synthesized *in vitro* were labeled with [α-^{32}P]GTP (400 Ci/mmol) or [γ-^{32}P]NTPs, synthesized as described (23), in a multiround transcription system (24, 25). Transcripts were analyzed on 6% polyacrylamide/7 M urea gels. *E. coli* RNA polymerase was a gift from Paul Haydock, Purdue University. *E. coli* RNA polymerase free of phosphatase (Boehringer Mannheim) was required for transcription using [γ-^{32}P]NTPs.

Enzyme Assays—Amidophosphoribosyltransferase (glutamine phosphoribosylpyrophosphate amidotransferase, EC 2.4.2.14) was assayed by the glutamate dehydrogenase method (26), β-galactosidase by the procedure of Miller (27), galactokinase as described by Wilson and Hogness (28), and β-lactamase by the method of Kelly *et al.* (29). Protein concentrations were determined by the method of Lowry *et al.* (30).

RESULTS

purF Control Region

The DNA sequence of *E. coli purF* and 463 bp of the 5′ flanking region has been previously reported (3). We have extended the DNA sequence an additional 420 bp in order to include the *purF* promoter and control region. The *purF* upstream region contains several potentially important elements: (i) an open reading frame capable of encoding a 162 amino acid polypeptide; (ii) a set of repetitive extragenic palindrome (REP) sequences (31) upstream of the *purF* promoter region; (iii) additional regions of dyad symmetry that are candidates for control signals. REP sequences are characteristic inverted repeats, about 35 nucleotides in length, that occur in nontranslated regions of a mRNA and may occupy up to 1% of the *E. coli* genome. The REP sequences and other dyad symmetries are indicated by *dots* in Fig. 1 and are *underlined* and numbered *I–VI* in Fig. 2. We have investigated the role of these elements in the expression and regulation of *purF*.

DNA Sequence—The nucleotide sequence upstream of *purF* was determined from both DNA strands using overlapping fragments and is shown in Fig. 2. We found a discrepancy with the DNA sequence previously reported (3). Bases C-105 and A-221 were omitted previously.

Analysis of the Upstream Open Reading Frame—As shown in Fig. 2, an open reading frame having the potential to code for a 162-amino acid protein is immediately upstream of the *purF* translation start site. Because there are no known genes linked to *purF*, we used *in vitro* transcription/translation to test whether the open reading frame encoded a protein. [^{14}C] Leucine-labeled proteins synthesized *in vitro* were resolved by sodium dodecyl sulfate-polyacrylamide gel electrophoresis and visualized by autoradiography. The results are shown in Fig. 3. Plasmid pSB2 (*purF*$^+$), *lane 1*, directed the synthesis of amidophosphoribosyltransferase, β-lactamase, and a protein with an estimated molecular weight of 15,000. Plasmid pCM3 (*purF*), *lane 2*, which contains 2.2 kb of *purF* upstream DNA subcloned into pUC8, also produced β-lactamase and the 15-kDa protein. Plasmid pCM3Δ170 (*purF*), *lane 3*, which contains a 170-bp *purF* promoter deletion in pCM3, no longer produced the 15-kDa protein. The calculated molecular weight from the predicted amino acid sequence of the upstream open

Downloaded from www.jbc.org by on November 9, 2006

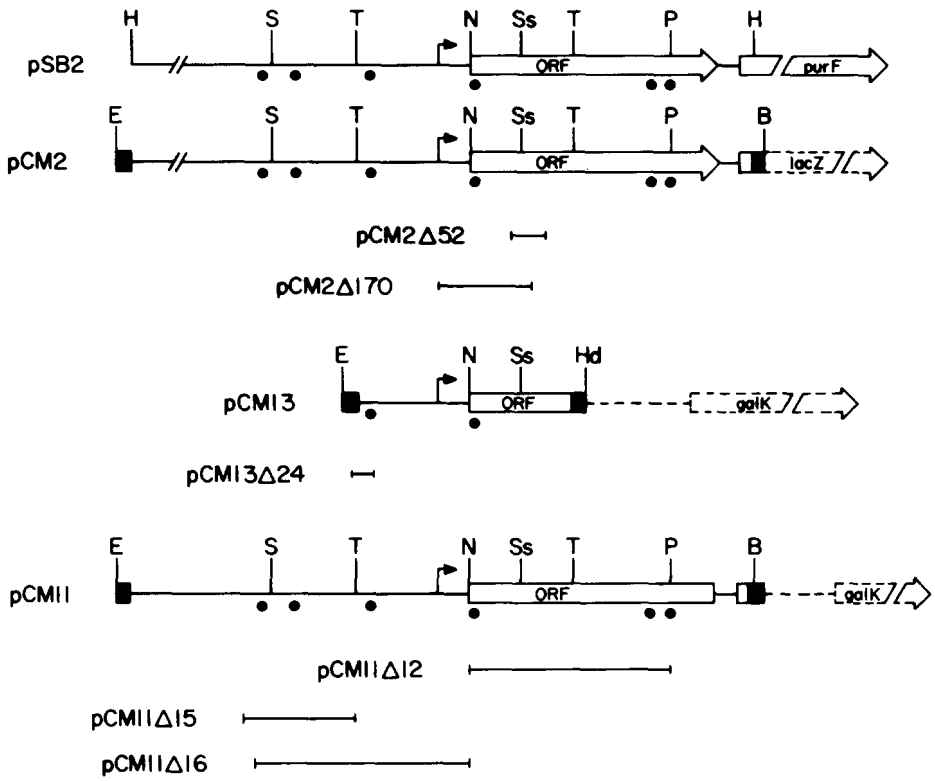

FIG. 1. **Maps of parental plasmids and deletions.** Only pertinent regions are shown. The *light line* denotes *purF* upstream DNA; *dashed line*, fusion vector DNA: *solid bars*, polylinker DNA: *open bars*, coding regions; *lines* with *brackets*, deletions made in the *purF* upstream region; *dots*, regions of dyad symmetry numbered I–VI in Fig. 2. The *in vivo* transcription start site (▶) is indicated as well as the direction of transcription. Abbreviations for restriction sites are: *B*, *Bam*HI; *E*, *Eco*RI; *H*, *Hinc*II; *Hd*, *Hin*dIII; *N*, *Nde*I; *P*, *Pvu*II; *S*, *Stu*I; *Ss*, *Sst*II; *T*, *Taq*I. *ORF*, open reading frame. Maps are not drawn to scale.

reading frame is 17,908. The experiment shown in Fig. 3 indicates that the protein of approximately 15 kDa synthesized from pSB2 and pCM3 is the 17.9-kDa product of the upstream open reading frame. We designate this product of the open reading frame protein 17.9.

To test the expression of the open reading frame *in vivo*, we attempted to construct *lacZ* fusions. Direct cloning of fragments that fuse the open reading frame with *lacZ* in the plasmid pMC1403 as well as deletion of DNA between the open reading frame and the *purF-lacZ* fusion in pCM2 were attempted. In-frame fusions with *lacZ* were not isolated but out of frame *lacZ* fusions were readily obtained. These results suggest that in-frame fusions of the open reading frame with *lacZ* were lethal. Fusions of *lacZ* with hydrophobic signal peptides that direct β-galactosidase into membranes are sometimes lethal (32, 33).

A notable feature of protein 17.9 is its high content of hydrophobic amino acids. Because of its deduced hydrophobic character and the apparent lethality of in-frame *lacZ* fusions, we evaluated the possibility that protein 17.9 has a membrane localization. The method of Kyte and Doolittle (34) was used to predict the hydropathy of protein 17.9. The hydropathy plot in Fig. 4 shows two regions in which the hydropathy index is greater than 1.0 for 20 or more amino acids. The mean hydropathy index for the two regions, residues 63–84 and 102–122, is 2.09 and 1.84, respectively. Such regions are predicted to span a membrane (34). This analysis suggests that protein 17.9 could be membrane bound.

Transcriptional Start Site—The transcriptional start site of *purF* was deduced by nuclease S1 mapping and *in vitro* transcription. The 5′ end of *purF* mRNA was first localized using a 1.5-kb *Hin*fI double-stranded probe. Fig. 5 shows that when the *Hin*fI fragment was used to probe RNA isolated from wild type strain W3110, a 400-bp protected fragment was detected (*lane 1*). This fragment was not obtained when tRNA was used as the source of RNA (*lane 2*). *Lane 3* shows the position of the untreated hybridization probe. The data in Fig. 5 thus localize the 5′ end of the mRNA to a site approximately 560 bp upstream of the *purF* coding region. This mRNA 5′ end corresponds to a site for transcription initiation designated P1.

To more precisely map the 5′ terminus of the *purF* mRNA a single-stranded *Taq*I fragment (−95 to 216) was used as a probe. Fig. 6 shows a single major protected fragment corresponding to a transcription start site at P1, 560 bp upstream of the *purF*-coding region. Transcription start site P1 is numbered +1 in Fig. 2.

To ensure that there is only one *in vivo* start site and that the P1 transcript extends into *purF*, a 498-base *Sst*II/*Hinc*II fragment (Fig. 1) was used as a probe. Fig. 7 shows that the coding strand of this probe (*lane 2*) but not the noncoding strand (*lane 1*) was fully protected, indicating that mRNA which initiates at +1 extends into *purF*.

The *purF* transcriptional start site was further characterized using *in vitro* transcription. Fig. 8 shows a schematic representation of four fragments used as templates and the run-off transcripts obtained from each. Fig. 9 shows the actual RNA run-off products obtained from *in vitro* transcription reactions using the four fragments as templates. Two major transcripts were produced from each fragment, and in all cases the two products were run-off transcripts originating at approximately +1 and +30. Further confirmation of these start sites was obtained using the four [γ-^{32}P]NTPs to label the 5′ end of the mRNA. *In vitro* transcription using each of the four [γ-^{32}P]NTPs indicated that only [γ-^{32}P]GTP was incorporated into transcripts originating at approximately +1 and +30 (data not shown). The first *in vitro* start site is assigned at G +1 and corresponds to P1, the start site observed *in vivo*. The second *in vitro* transcription start site, designated P2, is assigned to G +30 based on the size of the run-off transcript, spacing with a putative Pribnow box, TAAAAT (nucleotides 19–24), and labeling of the 5′ end of the transcript with [γ-^{32}P]GTP. G residues at positions +28 and +29 are not excluded as the start site for P2 but are less likely

Downloaded from www.jbc.org by on November 9, 2006

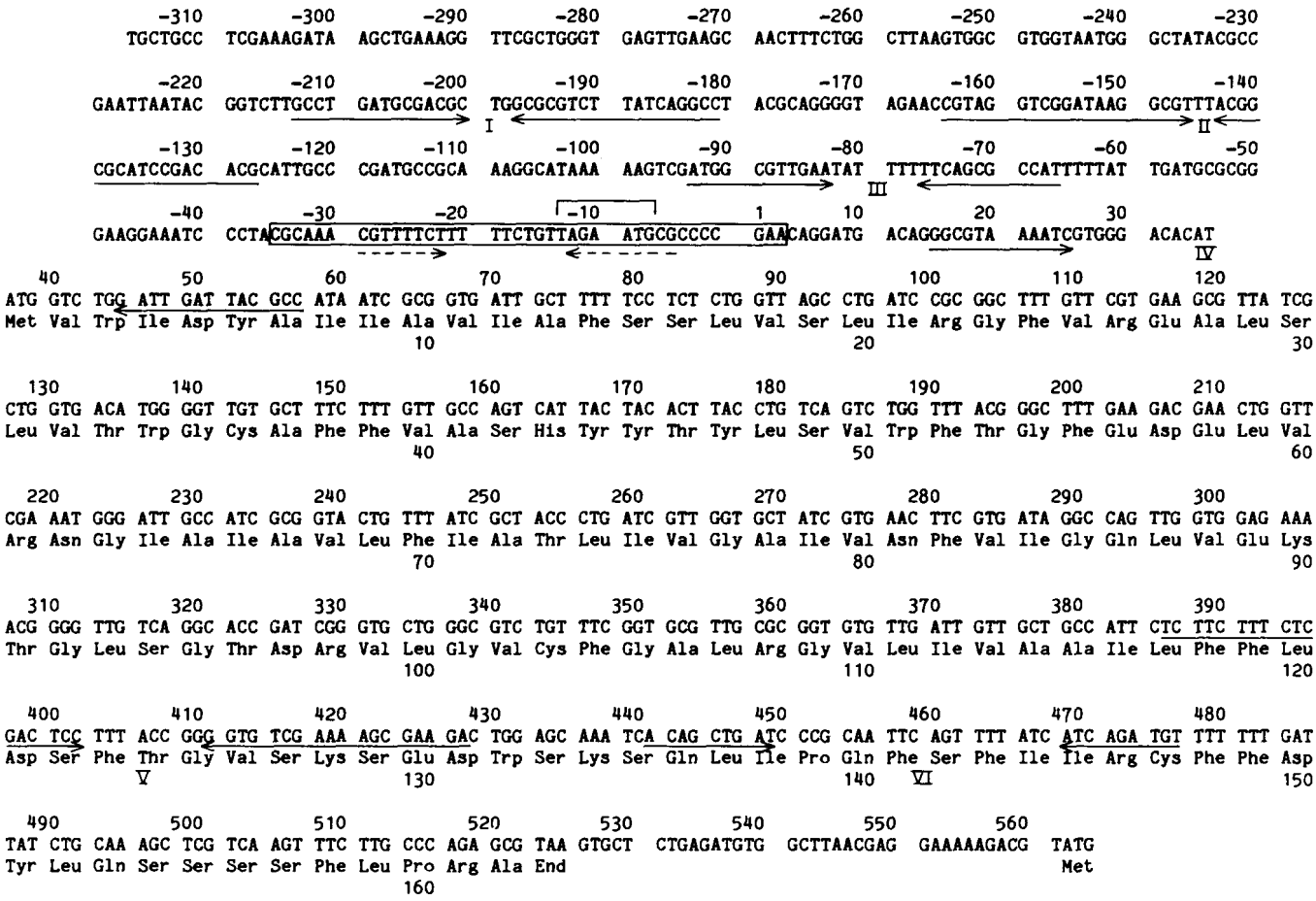

FIG. 2. **Nucleotide sequence and encoded polypeptide of *purF* upstream region.** Numbering of nucleotides is from the *in vivo* start of transcription. The Met codon, nucleotides 562–564, is the translation start for the *purF* coding region. Segments of dyad symmetry are marked by *arrows* and designated by *Roman numerals*. Dyad symmetries I and II are homologous to previously described REP sequences (31). The putative regulatory site is *boxed*, and an imperfect inverted repeat is identified by *dashed arrows*. The Pribnow box within the regulatory site is *overlined*.

based on their proximity to the putative Pribnow box at nucleotides 19–24. A transcript originating at P2 was not observed *in vivo*, under conditions which would have detected a signal approximately 1% as strong as that from P1.

Regulation of purF

Computer-assisted analysis of the DNA sequence surrounding the 5′ end of the *purF* mRNA revealed six regions of dyad symmetry that possibly could play a role in *purF* regulation. These six regions of symmetry are noted in Figs. 1 and 2. Calculated ΔG values for stem and loop formation, expressed as kcal/mol, are: I, −24; II, −23; III, −17; IV, −15; V, −17; and VI, −5. The next series of experiments was designed to determine whether regulation of *purF* occurs at the transcriptional level and to analyze what role, if any, dyad symmetries I–VI might have in *purF* regulation.

Measurement of purF mRNA by Hybridization—To determine whether adenine exerts repression at the level of transcription, *purF* mRNA levels were quantitated by RNA/DNA hybridization. RNA from strain TX158 bearing pSB2 was isolated from repressed and derepressed cells that had been pulse labeled with [^{3}H]uridine. RNA was hybridized to upstream and downstream DNA probes. These probes correspond to coding segments for protein 17.9 and amidophosphoribosyltransferase (Fig. 8, *top*). The data in Table II show 15–16-fold repression by adenine for both upstream and downstream mRNA segments. The stability of the two mRNA regions was also determined, and mRNA half-lives are shown in Table II. The half-life of downstream *purF* mRNA was approximately twice that of the upstream segment, independent of the repression status. Since the mRNA levels from repressed and derepressed cells were corrected for decay, the data in Table II reflect rates of mRNA synthesis. These data indicate that adenine repressed mRNA synthesis 15–16-fold.

Prior to pulse labeling of the cells a portion of each culture was removed for enzyme assay. Amidophosphoribosyltransferase activity was measured from cells grown either in the presence or absence of adenine. Table II shows that adenine caused an approximately 10-fold repression of amidophosphoribosyltransferase activity. This repression of *purF* enyme on a multicopy plasmid compares favorably with a value of 13-fold previously reported for a chromosomal *purF-lacZ* fusion (3, 36). β-Lactamase levels were measured and were found to be constant. Thus, changes in plasmid copy number cannot be a factor (data not shown). Repression of *purF* mRNA synthesis, therefore, accounts for the decrease in amidophosphoribosyltransferase enzyme levels.

Dyad symmetries V (nucleotides 386–428) and VI (nucleotides 441–477) bear similarity to the pause and attenuation signals observed in the *pyrBI* (37) and *pyrE* operons (38). The upstream and downstream probes used in the experiment summarized in Table II were chosen to investigate the possi-

Downloaded from www.jbc.org by on November 9, 2006

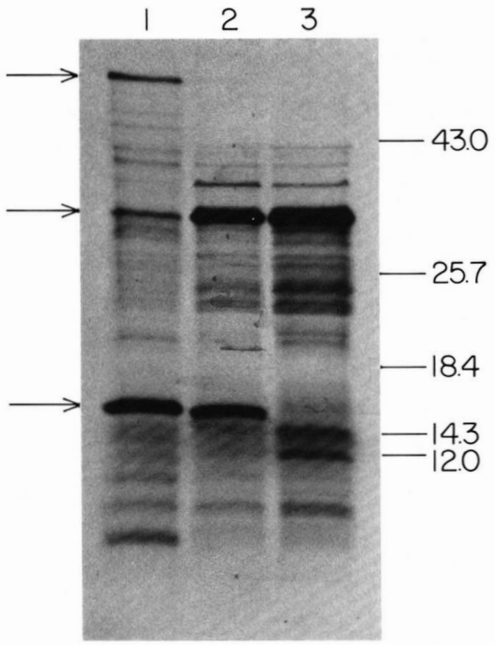

FIG. 3. **Synthesis of 17.9-kDa protein *in vitro*.** *Lane 1*, proteins synthesized from pSB2 (*purF*$^+$, ORF$^+$); *lane 2*, pCM3 (*purF*, ORF$^+$); and *lane 3*, pCM3Δ170, (*purF, ORF*$^-$). *Arrows* point to amidophosphoribosyltransferase, β-lactamase, and the 17.9-kDa product of ORF$^+$, *top* to *bottom*. ORF, open reading frame.

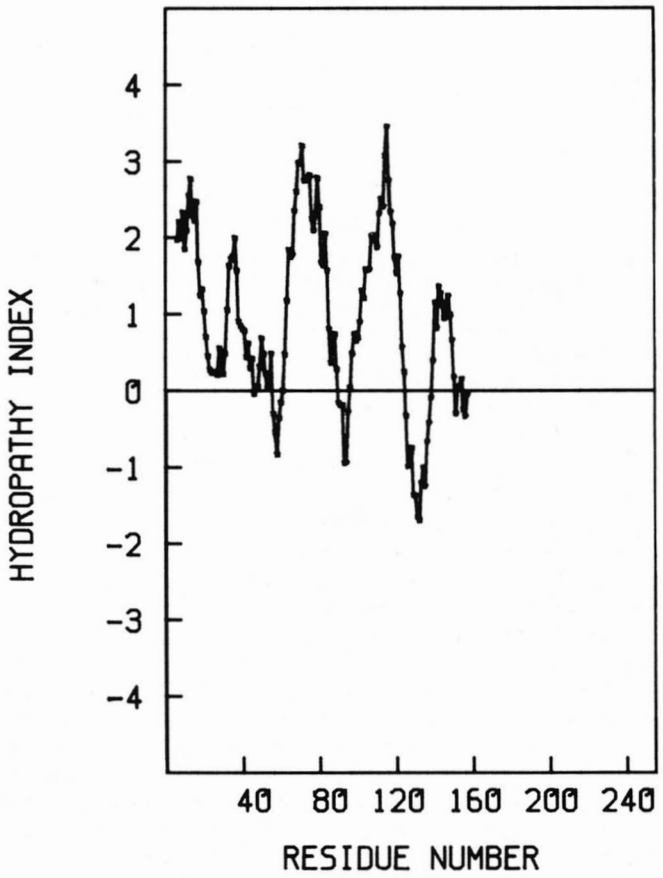

FIG. 4. **Plot of hydropathy index *versus* amino acid number.** Hydropathy was determined by the method of Kyte and Doolittle (34) using a span of 11 residues.

bility that attenuation is also involved in the regulation of *purF*. These probes hybridize to DNA upstream and downstream of dyad symmetries V and VI, respectively. The data

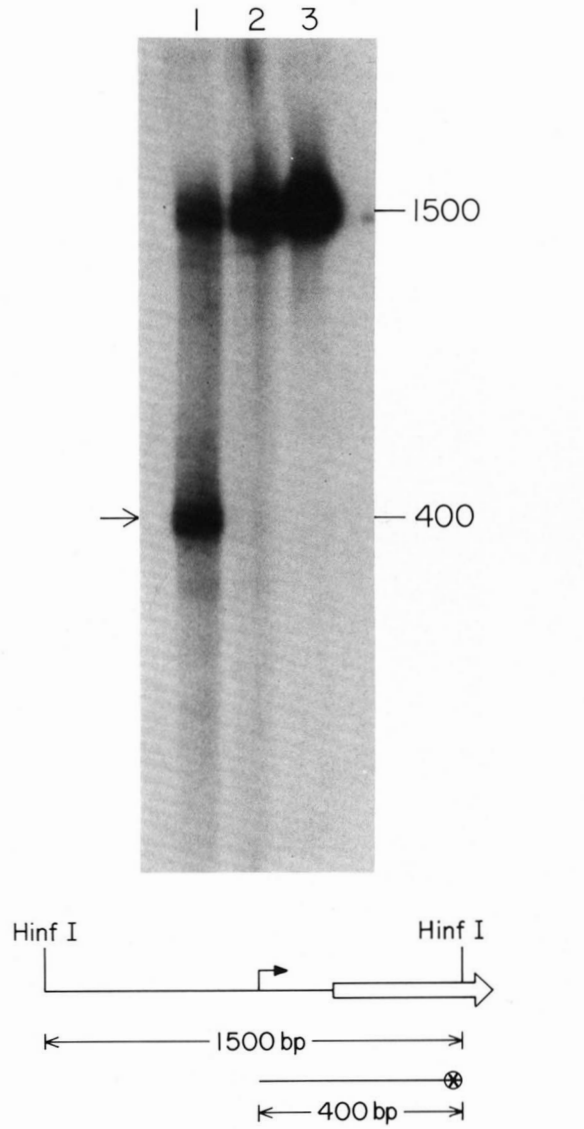

FIG. 5. **Localization of *purF* transcription start site.** A 1.5-kb double-stranded *Hin*fI fragment was used to probe RNA from wild type strain W3110 in a nuclease S1-mapping experiment. *Lane 1, E. coli* RNA; *lane 2*, tRNA; *lane 3*, undigested double-stranded *Hin*fI probe. An *arrow* marks the 400-bp RNA-protected fragment. The diagram shows the region used as the probe and the protected fragment. *, 5′ end labeled with ^{32}P; ▶, approximate transcription start site, ◊, coding region for protein 17.9.

in Table II show (i) equivalent levels of repression by adenine of upstream and downstream mRNA and (ii) an apparent 2-fold difference in the upstream *versus* downstream mRNA. The first result suggests that regulation by adenine does not cause attenuation of *purF* mRNA at dyad symmetries V and VI. However the significance of the apparent 2-fold difference is uncertain because we did not account for possible changes in the specific activity of the UTP pool during pulse labeling. The apparent 2-fold difference between upstream and downstream mRNA was not studied further because it was independent of repression by adenine and, therefore, has no regulatory significance.

Inspection of the DNA sequence around P1 indicates similarity to promoters that are under stringent control (39). P1 is similar to stringently controlled promoters in that it has a GC-rich sequence between the Pribnow box and the transcription start. To test for stringent control, cells were starved for

Downloaded from www.jbc.org by on November 9, 2006

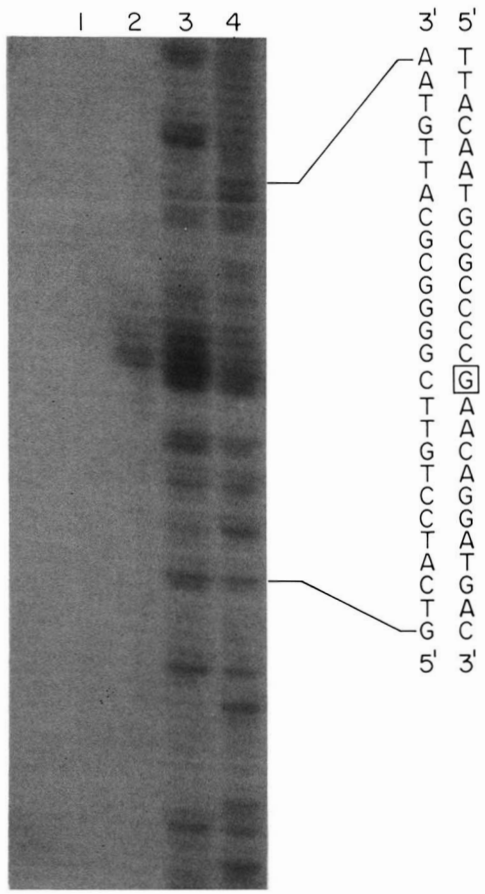

FIG. 6. **Identification of the 5′ end of the *purF* mRNA.** RNA from *E. coli* strain W3110 was hybridized to the single-stranded *Taq*311 probe labeled at the 5′ end. *Lanes 1* and *2* are the noncoding and coding strands, respectively, of the *Taq*311 fragment used as probes for S1 mapping. *Lanes 3* and *4* are G and AG sequencing ladders of the coding strand, respectively. A partial sequence from the sequencing ladder is shown, in which the transcription start site is *boxed* on the noncoding strand. The size of the protected fragment was adjusted by 1.5 nucleotides (35).

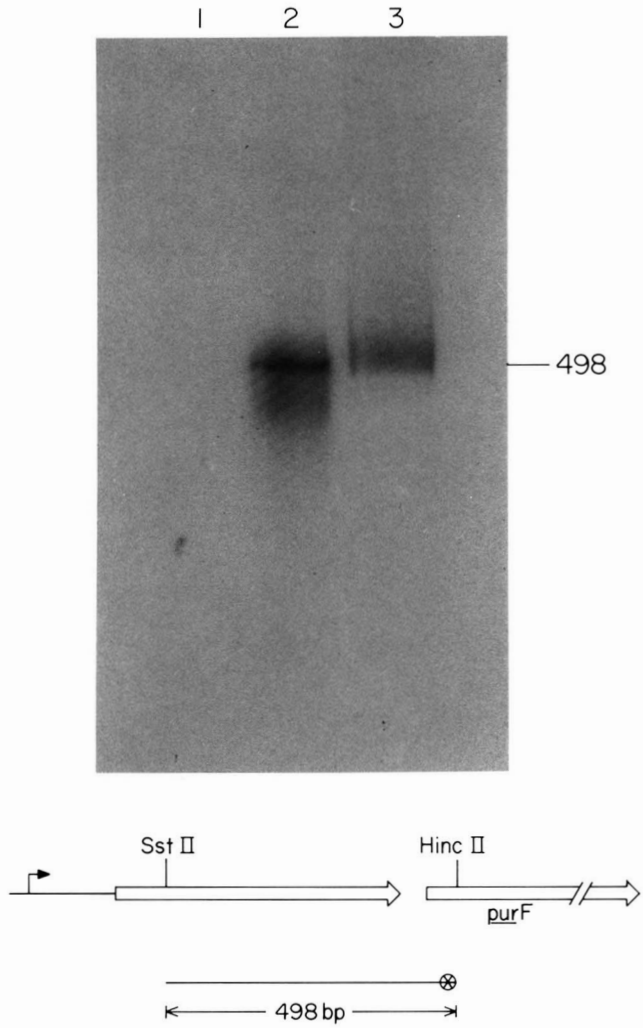

FIG. 7. **S1 mapping to demonstrate that the P1 transcript extends into *purF*.** Single-stranded *Sst*II/*Hinc*II fragments labeled at the 5′ termini were used to probe RNA from strain W3110. *Lanes 1* and *2* are the noncoding and coding strands used as probes for S1 mapping, respectively. *Lane 3* is the double-stranded *Sst*II/*Hinc*II fragment. The diagram shows the *Sst*II/*Hinc*II probe in relationship to *purF* and the transcription start site. *Symbols* are as in Fig. 5.

isoleucine by the addition of valine prior to pulse labeling. Starvation for isoleucine caused a reduction in the levels of *purF* mRNA. The data in Table II indicate that isoleucine starvation reduced mRNA levels by 25–58-fold. These data are uncorrected for mRNA decay and possible changes in UTP pool specific activity, and, therefore, comparisons between upstream and downstream mRNA are uncertain. The data in Table II provide initial evidence that *purF* transcription is subject to stringent control.

Deletion Analysis

A series of deletions was constructed in the DNA upstream of the *purF* coding region to define sequences involved in the regulation and expression of *purF* (Fig. 1). Both translational fusions to *lacZ* and transcriptional fusions to *galK* were used to monitor *purF* promoter activity. Cells harboring various derivatives of the *purF* control region fused to either *lacZ* or *galK* were grown in the presence or absence of a repressing level of adenine. Enzyme levels were measured and are shown in Table III. β-Lactamase levels were measured and used to correct for any variances in enzyme levels due to changes in plasmid copy number. Even though enzyme activities were corrected for changes in plasmid copy number, the repression factors obtained from *purF-lacZ* and *purF-galK* fusions were not comparable. We have observed that while the enzyme activities and repression factors obtained from a given construction were highly reproducible, repression factors varied between constructions. Maximal repression was obtained when *purF* was single copy on the chromosome. The smallest repression factors were obtained with multicopy *purF-galK* fusions. The reasons for this variation are not clear.

Dyad Symmetries I and II—Two regions of dyad symmetry that are homologous to previously described REP sequences (31) are located between nucleotides −213 and −182 and nucleotides −164 and −127. Plasmid pCM11Δ15 was constructed to determine whether these sequences have any role in the expression and regulation of *purF*. The deletion in pCM11Δ15 removes the DNA between nucleotides −209 and −96 and eliminates both copies of the REP sequence. Table III shows that plasmid pCM11Δ15 (REP^-) directs the same level of expression as plasmid pCM11 (REP^+), under both repressed and derepressed conditions. These results indicate that the two copies of REP sequence play no role in expression or regulation of *purF*.

Dyad Symmetries IV, V, and VI—A series of plasmids was constructed to investigate the role of dyad symmetries IV, V, and VI which are all located downstream of the *purF* pro-

Downloaded from www.jbc.org by on November 9, 2006

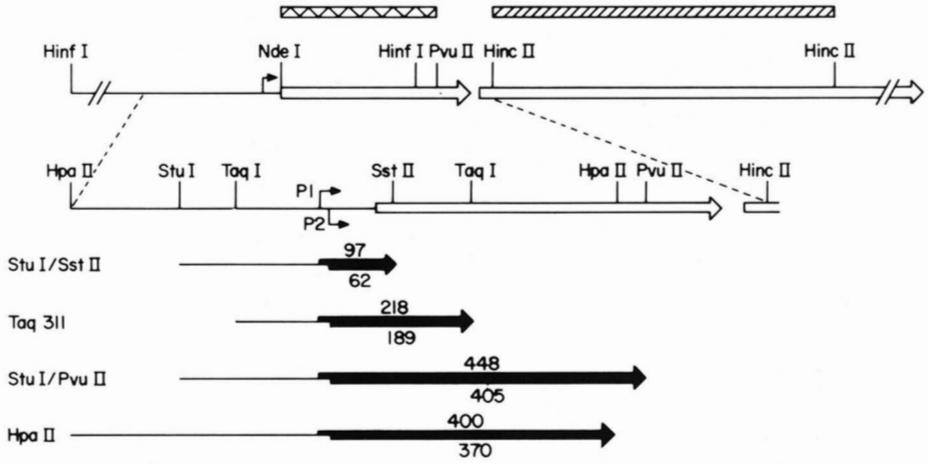

FIG. 8. **Schematic representation of fragments used and transcripts obtained in transcription experiments *in vitro*.** *Symbols* used are: *open arrows*, coding regions; *solid line* connected to *dark arrows*, restriction fragments used for *in vitro* transcription; *dark arrows*, run-off transcripts; ▨, DNA for upstream probe; ▨, DNA used for downsteam probe. P1 and P2 denote the two *in vitro* transcription start sites. The *numbers* indicate the sizes of run-off transcripts from P1 and P2.

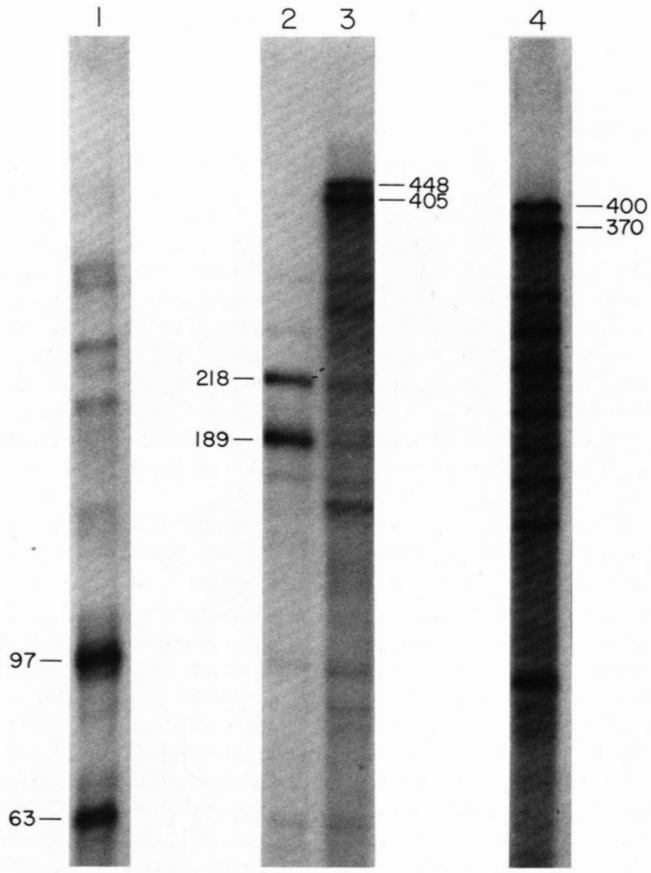

FIG. 9. **Run-off transcripts from *in vitro* transcription experiments.** Templates for *in vitro* transcription: *lane 1*, *Stu*I/*Sst*II; *lane 2*, *Taq*I-311; *lane 3*, *Stu*I/*Pvu*II; *lane 4*, *Hpa*II. Templates are shown schematically in Fig. 8.

TABLE II

Effect of growth conditions on purF mRNA and enzyme levels

DNA/RNA hybridization was performed as described under "Experimental Procedures." Hybridized mRNA was corrected for half-life and the number of uridine residues in the mRNA segment as described under "Experimental Procedures." Probes are shown schematically in Fig. 8.

Growth conditions	Probe	mRNA			Amidophosphoribosyltransferase	
		$t_{1/2}$	Hybridized	Repression	Activity	Repression
		sec	*cpm*	*-fold*	*units/mg protein*	*-fold*
Derepressed	Upstream	45	25,700			
	Downstream	107	13,100		0.74	
Repressed	Upstream	54	1,600	16		
	Downstream	139	864	15	0.07	10.5
Minimal	Upstream		3,590			
	Downstream		3,060			
Isoleucine-starved[a]	Upstream		137	25		
	Downstream		53	58		

[a] Cells were starved for isoleucine for 10 min prior to pulse labeling by the addition of valine (100 μg/ml).

moter. Plasmid pCM11Δ12 removes the DNA between nucleotides 35 and 444, corresponding to the right arm of symmetry IV, all of symmetry V, and part of the left arm of symmetry VI. Plasmid pCM11Δ12 shows similar levels of expression and regulation as pCM11. These results, therefore, rule out a role for dyad symmetries IV, V, and VI in regulation of *purF* by adenine.

Dyad Symmetry III—To investigate the role of dyad symmetry III in *purF* regulation, DNA between nucleotides −96 and −71 was deleted in pCM13. The resulting plasmid, pCM13Δ24, lacks dyad symmetry III. Although pCM13Δ24 exhibits only 31% of the promoter activity observed in the pCM13 parent, it is regulated normally. The reduced expression in pCM13Δ24 suggests that sequences upstream of −71 are necessary for high level *purF* expression. We have not eliminated the alternative possibility that vector sequences adjacent to −71 might have a negative effect on transcription. However, because pCM13 and pCM13Δ24 share common vector end points we favor the view that the reduced expression is a result of the deletion of promoter sequences and not a context effect. The normal regulation of pCM13Δ24 indicates that dyad symmetry III is not the *purF* operator and leads us to conclude that the *purF* regulatory region must lie between nucleotides −71 and 36.

Localization of the purF Promoter—The location of the *purF* promoter was confirmed by constructing deletions which remove the transcription start site and −10 region. Plasmid pCM2Δ170 contains a deletion from −27 to 177 while plasmid pCM11Δ16 has nucleotides −197 to 26 deleted. Both plasmids exhibit approximately 10% of the wild type derepressed promoter activity (Table III). The residual promoter activity in these plasmids is likely due to readthrough transcription from sites upstream of the *purF* operon and is no longer repressible by adenine. These results along with those obtained using plasmids pCM11Δ12 and pCM13Δ24 support our assignment

Downloaded from www.jbc.org by on November 9, 2006

TABLE III

Enzyme activities from wild type and derivative purF promoters

Strains CM1001 and N100 were hosts for *purF-lacZ* and *purF-galK* fusion plasmids, respectively. Enzymes were assayed in crude extracts made using a French press. β-Lactamase was assayed to determine changes in copy number. The plasmid copy number correction factor varied between 0.6 and 1.4. Strains CM1001/pMC1403 and N100/pK0100 were used to correct for non-*purF*-promoted expression (less than 5% of wild type derepressed levels).

	Intact symmetries	β-Galactosidase activity		Repression factor	Galactokinase activity		Repression factor
		Repressed[a]	Derepressed[b]		Repressed[a]	Derepressed[b]	
		units/mg protein			*units/mg protein*		
pCM2		121	1209	8.0			
pCM2X		131	1109	8.4			
pCM2Δ52		226	1183	5.2			
pCM2Δ170		127	104	0			
pCM11	I–VI				255	1014	4.0
pCM11Δ12	I–III				252	1022	4.0
pCM11Δ15	III–VI				258	941	3.6
pCM11Δ16	V–VI				124	132	0
pCM13	III, IV				178	605	3.4
pCM13Δ24	IV				61	189	3.1

[a] Grown in minimal medium plus 25 μg of adenine/ml.
[b] Grown in minimal medium.

of the *purF* promoter to the region between nucleotides −96 to +7.

Effect of Translation of Protein 17.9 on purF Expression—To determine if translational coupling (40, 41) plays a role in *purF* expression, two frame-shift mutations were constructed in the protein 17.9 coding region. Plasmid pCM2X contains *Xba*I linkers inserted at position 444 while pCM2Δ52 contains a 52-bp deletion at the 5′ terminus of the coding region. Both constructions cause premature translation termination in the protein 17.9 coding region. Table III shows that both plasmids direct approximately the same levels of *purF-lacZ* expression as wild type plasmid, pCM2. We, therefore, conclude that translational coupling does not play a role in *purF* expression and that an intact coding region for protein 17.9 has no role in *purF* expression in *cis*.

DISCUSSION

purF mRNA Is Polycistronic—The experiments reported in this paper provide the first analysis of the 5′ control region of a gene involved in *de novo* purine nucleotide synthesis. Combined data utilizing techniques for DNA sequencing, *in vivo* and *in vitro* mRNA mapping, and *in vitro* enzyme synthesis have established that *purF* is a component of a polycistronic mRNA. The *purF* coding region is preceded in the mRNA by an open reading frame that encodes a 17.9-kDa protein of unknown function. The following considerations lead to the conclusion that the 17.9-kDa protein, designated protein 17.9, could function as an integral membrane constituent and not as a component of the *purF* regulatory system. (i) The deduced amino acid sequence indicates that protein 17.9 is hydrophobic in nature. Kyte and Doolittle (34) have reported a good correlation between portions of proteins that interact with the cell membrane and their hydropathy. Integral membrane proteins generally contain stretches of 20 or more hydrophobic amino acids in which the average index of hydropathy is greater than +1.0 (34). Furthermore, membrane-spanning domains generally have a higher average index of hydropathy than hydrophobic regions found within a protein. A hydropathy prediction for protein 17.9 reveals stretches of 22 (residues 63–84) and 21 (residues 102–122) amino acids in which the index of hydropathy is comparable to that of known membrane-spanning domains (34). (ii) In-frame *lacZ* fusions to protein 17.9 are apparently lethal as are *lacZ* fusions to *lamB* and *malE* (32, 33). (iii) Insertions and deletions that disrupt the structure of protein 17.9 were without effect on the expression of *purF* or its repression by adenine. Insertion of *Xba*I linkers at nucleotide 444 changes the amino acid sequence after Leu-137 and should cause premature termination at position 138 while the deletion in pCM2Δ52 removes amino acids 13 to 29, changing the reading frame after residue 13, and should cause premature termination at position 132. Furthermore, in plasmid pCM13 residues after Val-60 of protein 17.9 were deleted, yet regulation by adenine was intact. These data exclude the possibility that protein 17.9 acts in *cis* for repression of *purF* or that the translational efficiency of *purF* is dependent upon ribosomes terminating translation of protein 17.9 at the normal UAA triplet, nucleotides 523–525.

The purF Promoter and Transcriptional Control—The 5′ end of the *purF* mRNA was determined by two independent methods, S1 nuclease mapping and *in vitro* transcription. The *in vivo* transcript, P1, starts 560 bp upstream of the *purF* structural gene. *In vitro* transcription experiments utilizing run-off transcription support the S1 nuclease-mapping data and establish that the 5′ end of transcript P1 is pppG. A second *in vitro* transcript, P2, originates approximately 30 bp downstream of P1. Transcription originating from P1 can utilize a good −10 region showing 6–7 identities with the concensus Pribnow box (42). Located upstream of P2 at nucleotides 19–24 is the sequence TAAAAT which might be utilized for transcription from the P2 site. Neither P1 nor P2 has a good candidate for a −35 region which maintains proper spacing.

Under the derepressed conditions used to isolate *E. coli* mRNA, P2 transcripts were not detected *in vivo*. If P2 were to function *in vivo*, it might be limited to situations where transcription from P1 was repressed by adenine or ppGpp. Further experiments are required to determine whether transcription from P2 can occur *in vivo*. The nucleotide sequence between the Pribnow box for P1 at positions −13 to −7 and the transcription start at G(+1) is identical to the concensus sequence in stable RNA promoters that are subject to stringent control (39). The data in Table II provide evidence that starvation for isoleucine decreases the level of *purF* mRNA as would be expected if stringent control of P1, the major promoter, were operative. Differential control of tandem promoters in a ribosomal operon has been demonstrated (43, 44).

As noted above, there are no good candidates for a −35

Downloaded from www.jbc.org by on November 9, 2006

promoter region 15–20 bp upstream of the Pribnow box for P1. The sequence information in Fig. 2 and the deletion analysis in Table III support the view that the *purF* promoter is one of a class of *E. coli* promoters that utilize upstream promoter sequences other than the −35 region. Lamond and Travers (45) have demonstrated that sequences between positions −40 and −98 are required for maximal transcription of *E. coli tyrT*. Two AT blocks, centered at −70 and −51, are contained within the *tyrT* promoter region. The block at −51 is also found in approximately the same position in most *rrn* P1 promoters. The *purF* upstream region contains two similar AT blocks centered at −79 and −63. Deletion 24 in plasmid pCM13 removes nucleotides −96 to −72 which contain the −79 AT block and causes a substantial decrease in promoter strength (Table III). The *purF* promoter is thus localized between nucleotides −96 and −7 and bears similarity to the *E. coli tyrT* (39, 45) and *rrn* P1 promoters (44, 45).

RNA/DNA hybridization experiments have established that *purF* is under transcriptional control (Table II). *purF* mRNA synthesis was repressed 15–16-fold by excess adenine while amidophosphoribosyltransferase levels were reduced 10.5-fold in the same cells.

REP Sequences—Dyad symmetries I and II exhibit homology to previously identified REP sequences (31). While a function has yet to be assigned to REP sequences, there is little evidence for a role in gene expression or regulation (31). Our deletion analysis indicates that the REP sequences located between nucleotides −213 and −127 have no effect on *purF* expression or regulation. Plasmid pCM11Δ15 (deletion of nucleotides −209 to −96) exhibits the same levels of expression and regulation as the wild type plasmid pCM11.

The presence of REP sequences between nucleotides −213 and −127 suggests that transcription of an upstream gene terminates nearby the *purF* promoter. REP sequences have thus far always been found within a transcribed but untranslated region of DNA, usually between two genes of an operon or at the 3′ end of a transcript. Several observations lend support to the idea that an upstream transcript terminates near the *purF* promoter. (i) Plasmids having 1.5 kb of DNA 5′ to *purF* produced an uncharacterized protein of M_r approximately 47,000 in a coupled transcription/translation system.[2] (ii) Higher levels of mRNA were detected by RNA/DNA hybridization using an upstream hybridization probe that was extended 148 bp from the *Nde*I site to the *Stu*I site[2] (Fig. 8). (iii) The deletions in pCM2Δ170 (nucleotides −27 to 144) and pCM11Δ16 (nucleotides −197 to 26) remove the *purF* Pribnow box and transcription start site but still allow between 8–13% of the wild type *purF* expression (Table III). In both cases the residual expression is not regulated by adenine. This nonregulated expression might result from readthrough from an upstream gene. Dyad symmetry III (nucleotides −93 to −66) is a likely candidate for a ρ-independent termination signal for the upstream transcript. Deletion analyses indicate that dyad symmetry III is not involved in regulation or expresson of *purF* (Table III).

The purF Control Region—Plasmid pCM13Δ24 which contains upstream DNA from nucleotides −71 to 216 exhibits low level *purF* expression and essentially normal regulation by adenine (Table III). DNA upstream of nucleotide −71 is likely required for *purF* promoter function but not regulation. DNA downstream of nucleotide 35 is not required for regulation as shown by the deletion in pCM11Δ12 which removes nucleotides 35–444 (Table III). The deletion analysis thereby localizes the regulatory region between positions −71 and +35. Recent experiments in the laboratory of John M. Smith[3] have identified a segment in the 5′ end of *purM* having 33 conserved bp with the 39 bp between −35 and +3 in the *purF* control region. This highly conserved region is *boxed* in Fig. 2. Located within this region is the imperfect inverted repeat sequence CGTTTTCT-N9-AGAATGCG which is characteristic of sites recognized by regulatory proteins (46). The overlapping position of the inverted repeat and −10 region suggests a mechanism for regulation. Binding of a regulatory protein to the inverted repeat sequence could block access of RNA polymerase to the Pribnow box and inhibit transcription initiation. The absence of a consensus −35 region and requirement for an upstream promoter element might result from the need to conserve the regulatory sequence between nucleotides −35 and +3. Other co-regulated *pur* regulon genes should also contain a homologous upstream control sequence.

Comparison of purF and pyr Gene Control Regions—There are several similarities between the *purF*, *pyrBI* (37), and *pyrE* (38) control regions. In all three operons there are tandem *in vitro* transcription start sites, an upstream open reading frame, and a set of dyad symmetries preceding the structural genes. In the case of *pyrBI* and *pyrE* attenuation is involved in regulation. Although dyad symmetries V and VI resemble the pause and termination elements found in *pyrBI* and *pyrE*, several results indicate attenuation does not function in *purF*. (i) Attenuated transcripts were not detected *in vitro*. (ii) Adenine-specific attenuation was not detected *in vivo* using RNA/DNA hybridization measurements. (iii) Constitutive *purF-lacZ* expression was observed from plasmids pCM2 and pCM2X when grown in the *purR* strain TX533.[2] Thus it appears that *purF* is not regulated by an attenuation mechanism but rather by repressor binding at regulatory sequences overlapping the promoter.

Acknowledgments—We thank John M. Smith, Louisiana State University Medical Center, for supplying *E. coli* strains and communicating the *purM-purF* sequence comparison. We are grateful to Randy Haun for help with the microcomputer, Kurt Brunden for the hydropathy analysis, and Ronald L. Somerville for helpful discussions and for reading the manuscript.

REFERENCES

1. Gots, J. S., Benson, C. E., Jochemsen, B., and Koduri, R. K. (1977) *Ciba Found. Symp.* **48,** 23–41
2. Koduri, R. K., and Gots, J. S. (1980) *J. Biol. Chem.* **255,** 9594–9598
3. Tso, J. Y., Zalkin, H., van Cleemput, M., Yanofsky, C., and Smith, J. M. (1982) *J. Biol. Chem.* **257,** 3525–3531
4. Makaroff, C. A., Zalkin, H., Switzer, R. L., and Vollmer, S. J. (1983) *J. Biol. Chem.* **258,** 10586–10593
5. Mäntsälä, P., and Zalkin, H. (1984) *J. Biol. Chem.* **259,** 8478–8484
6. McKenney, K., Shimatake, H., Court, P., Schmiessner, U., Brady, C., and Rosenberg, M. (1981) in *Gene Amplification and Analysis* (Chirikjian, J. G., and Papas, T., eds) Vol. 2, pp. 383–415, Elsevier, Amsterdam
7. Vieira, J., and Messing, J. (1982) *Gene* **19,** 259–268
8. Casadaban, M., and Cohen, S. N. (1980) *J. Mol. Biol.* **138,** 179–207
9. Vogel, H. T., and Bonner, D. M. (1956) *J. Biol. Chem.* **218,** 97–106
10. Casadaban, M. T., Chou, J., and Cohen, S. N. (1980) *J. Bacteriol.* **143,** 971–980
11. Messing, J. (1983) *Methods Enzymol.* **101,** 20–78
12. Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, pp. 135–139, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
13. Guo, L.-H., and Wu, R. (1983) *Methods Enzymol.* **100,** 60–96

[2] C. A. Makaroff, unpublished experiments.

[3] John M. Smith, unpublished data.

Downloaded from www.jbc.org by on November 9, 2006

14. Jinks-Robertson, S., Gourse, R. L., and Nomura, M. (1983) *Cell* **33,** 865–876
15. Berk, D. J., and Sharp, P. A. (1978) *Proc. Natl. Acad. Sci. U. S. A.* **75,** 1274–1278
16. Rose, J. K., Mosteller, R. D., and Yanofsky, C. (1970) *J. Mol. Biol.* **51,** 541–550
17. Gillespie, D., and Spiegelman, S. (1965) *J. Mol. Biol.* **12,** 829–842
18. Hauser, C. A., and Hatfield, G. W. (1984) *Proc. Natl. Acad. Sci. U. S. A.* **81,** 76–79
19. Zalkin, H., Yanofsky, C., and Squires, C. L. (1974) *J. Biol. Chem.* **249,** 465–475
20. Maxam, A. M., and Gilbert, W. (1980) *Methods Enzymol.* **65,** 499–560
21. Queen, C. L., and Korn, L. J. (1984) *Nucleic Acids Res.* **12,** 581–600
22. Sanger, F., Nickeln, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **80,** 3963–3965
23. Johnson, R. A., and Walseth, T. F. (1979) *Adv. Cyclic Nucleotide Res.* **10,** 135–167
24. Oxender, D. L., Zurawski, G., and Yanofsky, C. (1979) *Proc. Natl. Acad. Sci. U. S. A.* 5524–5528
25. Gardner, J. F. (1982) *J. Biol. Chem.* **257,** 3896–3904
26. Messenger, L. J., and Zalkin, H. (1979) *J. Biol. Chem.* **254,** 3382–3392
27. Miller, J. H. (1972) *Experiments in Molecular Genetics,* pp. 352–355, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
28. Wilson, D. B., and Hogness, D. S. (1966) *Methods Enzymol.* **8,** 229–240
29. Kelly, J. A., Frere, J. M., Duez, C., and Ghuysen, J. M. (1981) *Biochem. J.* **199,** 137–143
30. Layne, E. (1957) *Methods Enzymol.* **3,** 447–454
31. Stern, M. J., Ferro-Luzzi Ames, G., Smith, N. H., Robinson, E. C., and Higgens, C. (1984) *Cell* **37,** 1015–1026
32. Bassford, P. J., Silhavy, T., and Beckwith, J. R. (1979) *J. Bacteriol.* **139,** 19–31
33. Emr, S. D., Hall, M. N., and Silhavy, T. J. (1980) *J. Cell Biol.* **86,** 701–711
34. Kyte, J., and Doolittle, R. F. (1982) *J. Mol. Biol.* **157,** 105–132
35. Sollner-Webb, B., and Reeder, R. H. (1979) *Cell* **18,** 485–499
36. Smith, J. M., and Gots, J. S. (1980) *J. Bacteriol.* **143,** 1156–1164
37. Turnbough, C. L., Hicks, K. L., and Donahue, J. P. (1983) *Proc. Natl. Acad. Sci. U. S. A.* **80,** 368–372
38. Poulsen, P., Bonekamp, F., and Jenson, K. F. (1984) *EMBO J.* **3,** 1783–1790
39. Travers, A. A. (1984) *Nucleic Acids Res.* **12,** 2605–2618
40. Oppenheim, D., and Yanofsky, C. (1980) *Genetics* **95,** 785–795
41. Schumperli, D., McKenney, K., Sobieski, D. A., and Rosenberg, M. (1982) *Cell* **30,** 865–871
42. Rosenberg, M., and Court, P. (1979) *Annu. Rev. Genet.* **13,** 319–353
43. Sarmientos, P., Sylvester, J. E., Contente, S., and Cashel, M. (1983) *Cell* **32,** 1337–1346
44. Sarmientos, P., and Cashel, M. (1983) *Proc. Natl. Acad. Sci. U. S. A.* **80,** 7010–7013
45. Lamond, A. I., and Travers, A. A. (1983) *Nature* **305,** 248–250
46. Giequel-Sanzey, B., and Cossart, P. (1982) *EMBO J.* **1,** 591–595

The Journal of Biological Chemistry

Downloaded from www.jbc.org by on November 9, 2006