Cloning and Sequence of *Bacillus subtilis* *purA* and *guaA*, Involved in the Conversion of IMP to AMP and GMP†

PEKKA MÄNTSÄLÄ‡ AND HOWARD ZALKIN*

Department of Biochemistry, Purdue University, West Lafayette, Indiana 47907-1153

Received 12 December 1991/Accepted 17 January 1992

***Bacillus subtilis* genes *purA*, encoding adenylosuccinate synthetase, and *guaA*, coding for GMP synthetase, appear to be lethal when cloned in multicopy plasmids in *Escherichia coli*. The nucleotide sequences of *purA* and *guaA* were determined from a series of gene fragments isolated by polymerase chain reaction amplification, library screening, and plasmid rescue techniques. Identifications were based on amino acid sequence alignments with enzymes from other organisms. Comparison of the 5'-flanking regions of *purA* and *guaA* with the *pur* operon suggests similarities in mechanisms for gene regulation. Nucleotide sequences are now available for all genes involved in the 14-step pathway for de novo purine nucleotide synthesis in *B. subtilis*.**

The de novo synthesis of AMP and GMP proceeds by a 14-step branched pathway via IMP (see outline in Fig. 1). A connection between pathways for synthesis of purine nucleotides and histidine is also shown in Fig. 1. Although the de novo pathway for synthesis of purine nucleotides is invariant, the genetic organization and regulation of expression differ between organisms. In *Bacillus subtilis*, genes encoding enzymes for 10 steps to IMP are in a 12-gene *pur* operon (3). Gene *purB* in the *pur* operon encodes an enzyme catalyzing step 8 in the pathway to IMP and also the final reaction in the branch to AMP. The 12-gene *pur* operon has been cloned, the genes have been sequenced, and initial studies on expression have been reported (3–6). The *pur* operon is subject to dual control by adenine and guanine compounds. An adenine compound represses transcription initiation and a guanine compound regulates transcription termination-antitermination in an mRNA leader region preceding the first structural gene. *B. subtilis* gene *guaB* has been cloned (16) and sequenced (11), but there is no information on mechanisms for regulation. Earlier studies had shown that the levels of some of the enzymes involved in the synthesis of IMP are repressed by the addition of purines to the medium (17, 18). More recently, levels of enzymes encoded by three genes of the *pur* operon as well as *purA* and *guaB* were shown to respond to addition of purine compounds to the growth medium (22).

The only genes for the de novo pathway that have not been isolated from *B. subtilis* are *purA* and *guaA*. Here, we report the cloning and sequence determination of *purA* and *guaA*, along with identification of promoter regions. In *guaA* there is a 130-nucleotide 5' untranslated region that has the capacity for secondary structure, similar but not identical to that in the *pur* operon.

MATERIALS AND METHODS

Bacterial strains, plasmids, and media. Strains and plasmids are described in Table 1. LB medium (9) was used for growth of *Escherichia coli*. Transformants were selected on

LB agar with 50 µg of ampicillin per ml. Minimal medium for growth of *E. coli* and *B. subtilis* was M9 salts (15) supplemented with 0.5% glucose, 0.2% casein hydrolysate, and 1 µg of thiamine per ml. *B. subtilis* transformants were selected on Penassay broth agar (Difco) with 5 µg of chloramphenicol per ml. Procedures for transformation of *E. coli* (13) and *B. subtilis* (1) have been described previously.

Amplification of *purA* and *guaA*. The polymerase chain reaction (PCR) was used to amplify fragments of *purA* and *guaA* from *B. subtilis* chromosomal DNA (31). DNA from *B. subtilis* DE1 was prepared by the method of Saito and Miura (19). Reaction mixtures of 50 µl contained 40 ng of DNA, 200 ng of each oligonucleotide primer (see Fig. 2), 25 mM Tris-HCl (pH 8.5), 50 mM KCl, 3 mM MgCl₂, 0.5 mM each nucleoside triphosphate (NTP), and 2.5 U of *Thermus aquaticus* DNA polymerase (Perkin-Elmer Cetus). Template DNA was denatured at 92°C for 1 min, annealed at 50°C for 2 min, and extended for 2 min at 72°C for 37 cycles. PCR products were extracted with phenol, precipitated with ethanol, and

FIG. 1. Outline of the pathway for de novo purine nucleotide synthesis. The intermediate AICAR, 5-phosphoribosyl-4-carboxamide-5-aminoimidazole, is also a product of the histidine pathway. Gene-enzyme relationships are as follows: *purA*, adenylosuccinate synthetase; *purB*, adenylosuccinate lyase; *guaB*, IMP dehydrogenase; *guaA*, GMP synthetase; *guaC*, GMP reductase.

* Corresponding author.

† Journal paper 13281 from the Purdue University Agricultural Experiment Station.

‡ Permanent address: Department of Biochemistry, University of Turku, SF-20500 Turku 50, Finland.

TABLE 1. Strains and plasmids used

Strain or plasmid	Description	Source or reference
<i>Bacillus subtilis</i> DE1	Prototrophic revertant of W168 <i>trpC2</i>	3
<i>Escherichia coli</i>		
JM101	$\Delta(lac-pro)$ <i>thi supE F' traD36 proAB lacI^qZAM15</i>	14
DH5 α F'	F' $\phi 80$ <i>dlacZAM15</i> $\Delta(lacZYA-argF)$ U169 <i>deoR recA1 endA1 hsdR17</i> ($r_K^- m_K^+$) <i>supE44</i> $\lambda^- thi-1 gyrA96 relA1$	9
KE94	F' <i>hsdS20</i> ($r_B^- m_B^-$) <i>recA13 ara-14 proA2 lacY1 galK12 rpsL20</i> (<i>str</i>) <i>xyl-5 mtl-1</i> <i>supE44</i> $\lambda^- pcnB80 zad::Tn10$; used to maintain low plasmid copy number	12
Plasmids		
pDE194	pUC8 derivative that contains pC194-CAT	3
pMSKY3	pUC18 derivative that contains <i>spac</i> promoter in front of CAT-86	3
pWSK29	Low-copy-number plasmid	28
pPP11	pDE194 derivative with a 324-bp <i>purA</i> HindIII fragment (198–521)	
pPP12	pMSKY3 derivative with a 632-bp <i>purA</i> EcoRI fragment (478–1109)	
pPG13	pMSKY3 derivative with an 834-bp <i>guaA</i> EcoRI fragment (216–1049)	
pPG14	pDE194 derivative with a 471-bp <i>guaA</i> PvuII fragment (399–869)	
pPPA11	<i>purA</i> 5' region (–(–)6000–521) rescued by <i>Bam</i> HI in pPP11	
pPPA12	<i>purA</i> 3' region (478–5000) rescued by <i>Pst</i> I in pPP12	
pPGA13	<i>guaA</i> 3' region (216–2250) rescued by <i>Pst</i> I in pPG13	
pPGA14	<i>guaA</i> 5' region (–(–)1200–869) rescued by <i>Nco</i> I in pPG14	
pPPA15	Subcloned 663-bp <i>purA</i> EcoRI fragment ((–)178–477) in pWSK29	
pPGA16	Subcloned 780-bp <i>guaA</i> EcoRI fragment (1050–1829)	

digested with enzymes appropriate for the adapter sites. Electrophoresis was in gels of 3% Nusieve GTG low-melting-temperature agarose (FMC Bioproducts). Amplified DNA fragments were isolated from gel slices as described previously (20).

Screening of *B. subtilis* genomic libraries. PCR fragments were labeled by random priming (7) or by primer extension in phagemid pUC118 with 1 μ g of single-stranded DNA template, 1.8 ng of universal primer, 4 U of Klenow fragment, 25 μ M dNTPs except dCTP, which was 40 μ Ci of [32 P]dCTP (3,000 Ci/mmol), and buffer in a volume of 30 μ l (20). The fragments were labeled for 10 min at 25°C and then chased for 10 min with 25 μ M dCTP. Plasmid libraries prepared from *B. subtilis* DNA digested with HindIII, EcoRI, PstI, SphI, and HincII were transformed into *E. coli* DH5 α F' cells and plated to give 500 to 2,000 colonies per plate. Nitrocellulose filters were prepared according to the supplier (Schleicher & Schuell) and hybridized with 32 P-labeled PCR probe as described previously (31).

Plasmid rescue. *B. subtilis purA* and *guaA* DNA fragments were ligated into vectors pDE194 and pMSKY3 for integration and rescue. Plasmid miniscreen DNA (1 μ g) was used to transform 0.1 ml of frozen competent *B. subtilis* cells (1). After shaking for 30 min at 37°C, samples were plated on Penassay broth agar plates containing 5 μ g of chloramphenicol per ml. About 1.5 μ g of chromosomal DNA from a transformant was digested with an appropriate restriction enzyme, religated, and transformed into *E. coli* DH5 α F'.

DNA sequencing. Nucleotide sequences were determined by the chain termination method (21) modified for Sequenase polymerase (U.S. Biochemicals) (24). In some cases, DNA fragments were subcloned into pUC118 and pUC119 to obtain single-stranded phagemid templates. The nucleotide sequence of the 5'- and 3'-flanking regions of *purA* and the 5'-flanking region of *guaA* was determined from double-stranded DNA purified by the alkaline lysis method (2). DNA was denatured with alkali, neutralized, precipitated, and sequenced (24).

Primer extension mapping. Cells were grown in 200 ml of minimal medium (31) to Klett 70 with a 66 filter. RNA was

isolated by a hot-phenol extraction method (4, 23). The yield of RNA was approximately 400 μ g with an A_{260}/A_{280} ratio of about 2.05. Primers for *purA* were a 19-mer complementary to nucleotides 99 to 117 and a 17-mer complementary to

FIG. 2. Strategy for PCR amplification. (A) Two conserved regions of adenylosuccinate synthetase from *E. coli* and mouse were amplified by using degenerate primers: left, 5'-CAA/GTGGGNNGA T/CGAA/GGG; right, 5'-ACNACNCCG/ATTNCCA/G/TAT. (B) Two conserved regions in the glutamine amide transfer domain in *E. coli* GMP synthetase (GMPs). Degenerate primers were designed to this region in *E. coli* GMP synthetase and four other *E. coli* glutamine amidotransferases as well as formylglycinamide ribonucleotide synthetase from *B. subtilis* (3): left, 5'-GGNA/GTNTGT/CTAT/CGG NATGCA; right, 5'-T/CTCNGGA/GTGA/GAAT/CTGNACNC. (C) Strategy for amplification of *purA* fragment 3 (see Fig. 3A). Primers were designed to a sequence in fragment 2 (Fig. 3A) and to a conserved region in *E. coli* and mouse GMP synthetase: left, 5'-AGTGGAAGAAGAGCG; right, 5'-ANCCNGGCATNGTT/CTCA/GTA. Primers in panels A and B contained *Eco*RI and *Hind*III adapters; primers in panel C contained *Sph*I and *Bam*HI adapters. Adapters are shown by a bent line.

FIG. 3. Schematic diagram of cloning procedures. The thick line is the protein-coding region. Numbering is from the transcription start site (+1). Relevant restriction enzyme cleavage sites used for cloning are B, *Bam*HI; E, *Eco*RI; H, *Hind*III; N, *Nco*I; P, *Pst*I; Pv, *Pvu*II. Numbered lines show the length and position of cloned fragments. (A) Cloning of *purA*. The fragments are as follows: 1, primary PCR isolate; 2, screened from *Hind*III library (pPP11); 3, PCR isolate; 4, plasmid rescue (pPPA12); 5, plasmid rescue (pPPA11); 6, subcloned 663-bp *Eco*RI fragment in pPPA15. (B) Cloning of *guaA*. Fragments are as follows: 1, primary PCR isolate; 2, screened from *Eco*RI library (pPG13); 3, plasmid rescue (pPGA13); 4, plasmid rescue (pPGA14).

nucleotides 132 to 148. Primers for *guaA* were an 18-mer complementary to nucleotides 95 to 112 and a 19-mer complementary to nucleotides 157 to 175. Primers were labeled with [γ - 32 P]ATP by using polynucleotide kinase. About 10 ng (400,000 cpm) of labeled primer was annealed to 25 ng of RNA in a solution containing 10 mM Tris-HCl (pH 8.0), 1 mM EDTA, and 10 U of RNasin RNase inhibitor at temperatures of 90, 65, and 42°C for 5 min each. A solution containing 50 mM Tris-HCl (pH 8.3), 140 mM KCl, 10 mM MgCl₂, 5 μ g of nuclease-free bovine serum albumin, 1 mM dithiothreitol, a mixture of the four dNTPs (0.5 mM each), and 18 U of avian myeloblastosis virus reverse transcriptase was added, and incubation was performed for 2 h at 42°C. The cDNA products were precipitated with ethanol and electrophoresed in a 6% polyacrylamide gel along with a sequencing ladder.

DNA blots. DNA (1 μ g), digested with appropriate restriction enzymes, was size fractionated on a 1% agarose gel, blotted to Nytran (Schleicher & Schuell) nylon membrane, and hybridized with 32 P-labeled probe.

Nucleotide sequence accession number. The sequences of *purA* and *guaA* have been deposited in the GenBank data base under accession numbers M83690 and M83691, respectively.

RESULTS

Cloning of *purA*. We were unable to isolate *purA* or *guaA* from low-copy (10) or multicopy plasmid libraries of *B. subtilis* DNA by complementation of *E. coli purA* or *guaA* purine auxotrophs. Accordingly, we turned to an approach based on PCR amplification of an internal segment of each gene and then hybridization screening and plasmid rescue. PCR amplification of an internal region of *purA* conserved in *E. coli* (30) and mouse (8) DNA is diagrammed in Fig. 2A. An amplified product of the expected size, about 200 bp, was isolated from an agarose gel. For ligation into a plasmid vector, it was necessary to digest the adapter sites with *Eco*RI and *Hind*III. As a result of an internal *Hind*III site,

fragments of about 130 and 70 bp were recovered and cloned into pUC119. By nucleotide sequence analysis, 68- and 126-bp subfragments of a 194-bp sequence with an internal *Hind*III site were identified which were similar but not identical to a fragment of *E. coli purA*. Southern analysis confirmed that the PCR product was derived from *B. subtilis* DNA. This fragment is designated 1a-1b in Fig. 3A.

The *purA* gene sequence was reconstructed from four additional clones as outlined in Fig. 3A. (i) Fragment 2, a 324-bp *Hind*III fragment, was isolated by screening a plasmid library with fragment 1b as the hybridization probe. (ii) Fragment 3, 785 bp, was isolated by PCR amplification with primers from sequence in fragment 2 and a conserved region at amino acids 373 to 379 in *E. coli* and 410 to 416 in the mouse sequence (Fig. 2C). (iii) Integration of a 632-bp *Eco*RI segment of fragment 3 in plasmid pPP12 was used to rescue a 4.5-kb 3' segment of *purA* DNA in pPPA12, fragment 4. Transformants were obtained only when the plasmid copy number was suppressed in strain KE94. (iv) Integration of fragment 2 in pPP11 was used to rescue a 6-kb *purA* 5'-flanking region in pPPA11, fragment 5. Most attempts to subclone smaller *purA* fragments from pPPA11 were unsuccessful. However, a 663-bp *Eco*RI fragment was isolated in a low-copy-number vector, designated pPPA15. The 663-bp *Eco*RI fragment in pPPA15 is labeled fragment 6 in Fig. 3A. Southern blots of chromosomal DNA with cloned fragments as probes were consistent with a single gene having the restriction sites shown in Fig. 3A.

Cloning of *guaA*. A similar strategy starting with PCR amplification from *B. subtilis* chromosomal DNA was used to clone *guaA*. PCR primers were designed to two regions of *E. coli guaA* (26) corresponding to conserved sequences in the glutamine amide transfer domain of six glutamine amidotransferases (3). The expected size of a PCR product is about 310 bp including the adapters (Fig. 2B). Three bands of approximately 290, 320, and 350 bp were detected after electrophoresis. A putative *guaA* fragment of 278 bp, fragment 1 in Fig. 3B, was identified by sequence analysis of a clone containing DNA from the smallest of the three bands.

-121 TATGACAAAGCAGTGAAGAAACGCTTATAAAACGAATGGAAGCGAACGAATATAGATTTA
 -35 -10
 -61 CAATAAATTAATGTTCCGATTACAAATGACTTCTGTCTCTCACTGATTAACCTTGATT
 +1 M S S V V
 -1 TGTTTGAATAGATCGTTTGAAGGTTAAACGAGGTCACGACATGCTCTCAGTAGTT
 6 V V G T G W G D E G K G K I T D F L S E
 60 GTAGTAGGTACGCAATGGGCGATGAAGAAAAAGTAAATTAACAGATTCTCATCAGAA
 26 N A E V I A R Y Q G G N N A G H T I K F
 120 AATGCAGAAATGATCGCCGTTATCAAGCGGAAATAACGAGGCGATACATCAAGTTT
 46 D G I T Y K L H L I P S G I F Y K D K T
 180 GACGGAATCACATATAAGCTTCACTTAATCCCGTCTGGAATTTCTATAAGGATAAAACG
 66 C V I G N G M V V D P K A L V T E L A Y
 240 TGTGTAATCGGAAACGGAATGGTTGTAGATCCGAAAGCATTAGTCACAGAGCTTCGCTAT
 86 L H E R N V S T D N L R I S N R A H V I
 300 CTTCTAGCGCAACGTGAGTACAGATAAAGTGAATGAGACAGACAGAGCTCACGCTCATT
 106 L P Y H L K L E D V E E E R K G A N K I
 360 CTGCGGTATCATTTGAAATGGATGAAGTGAAGAAAGCGTAAAGGGGCTAAACAGATC
 126 G T T K K G I G P A Y M D K A A R I G I
 420 GGCACAAAGAAAAAGGAATCGGCCCTGCTTACATGGATAAAGCAGCCGATCGGAATT
 460 R I A D L D L Q G T F A K A L V T E L A Y
 480 CGCATCGCGGATCTGTAGACCGTGACGCTTTGCGGAAAAGCTTGAGCGCAATCTTGAA
 166 E K N R L L E K M Y E T E G F K L E D I
 540 GAAAAACCGCTCTTCGAGAAAAATGTCAGACAGAAAGGGTTTAAACTTGAGGATATC
 186 L D E Y Y E G V Q I K K Y V C D T S V
 600 TTAGACGAATATTATGAGTACGACAGCAGATTAAGATGATGTTTCGATACATCTGTT
 206 V L N D A L D E G R R V L F E G A Q G V
 660 GTCTTAAACGATCTCTTATGATGAAGGGCGCGCTGATATTATTTGAAGGCGCAAGGGGTT
 226 M L D I D Q G T Y P F V T S S N P V A A
 720 ATGCTCGATATCGACCAAGGAACATACCGTTGTTGTCATCTAACCGGTTGCGCGG
 246 V S Q S V L V S A R P K I K H V V G V S
 780 GTGTCACAGTCGGTTCGGTGTGCGGCCGACCAAAAAATCAAGCAGTGTGTCGGTGTATCA
 266 K A Y T T R V D G P F P V T E D L K D E I
 840 AAAGCATATACGACTCGTGTGCGGCGCGTCTTTTCCGACTGAGTCAAGGATGAATC
 286 G D Q I R E V G R E Y G T T T G R P P R
 900 GGGCATCAATCCGTGAAGTCGGACGCGAATGGAACAACAACAGGCGCGCGCGCGGT
 306 V G W F D S V F V R H A R R V S G I T D
 960 GTGCGGTGGTTTGACAGCGTTGTGTGCGCCACGCGCGCGTGTAGCGGAATTCAGAT
 326 L S L N S I D V L A G I E T L K I C V R
 1020 CTTTCTCTGAACCAATTGACGCTCTAGCAGGAATGAAACGTTGAAATCTGTGTGCGG
 346 Y R Y K G E I I E E F P A S L K A L A E
 1080 TACCGCTACAAAGCGAAATCATTTGAAGAAATCCAGCAAGTCTTAAGGCACTTGTCTGAA
 366 C E P V Y E E M P G W T E D I T G A K S
 1140 TGTGAGCGGTATATGAAGAAATGCGCGGCTGAGTCTGAGGATATTACAGGTGCGAAGAGC
 386 L S E L P E N A R H Y L E R V S Q L T G
 1200 TTGAGCGAGCTTCCGGAATGCGCGCAATTATCTTGAAGCGTGTGTCTGAGCTGACAGGC
 406 I P L S I F S V G P D R S Q T N V L R S
 1260 ATTCCGCTTCTATTTCTCTGTGCGTCCAGACCGCTCACAACAAATGCTCTTCGCAAT
 426 V Y R A N
 1320 GTGTACCGTGGAACTAAATAGAATATCGTCTGCAAGCCCTTATTTAAGGGGCTGTGTTT
 1380 TTTGGTTGTTTGAACGCGCGCATATAAGTTGGTCTGAGAAAAATATGAAAAACCAAA
 1440 AAATTCGCTTCTTATTTTCTCTGTCGCTGCGTCCAGACCGCTCACAACAAATGCTCTTCGCA

FIG. 4. Nucleotide sequence of *purA* and derived amino acid sequence of adenylosuccinate synthetase. Numbering of the nucleotide sequence is from the start of transcription. Possible -35 and -10 promoter elements and a ribosome binding site (Shine-Dalgarno [SD]) are marked. A potential secondary structure for rho-independent transcription termination is marked in the 3'-flanking region. Although not shown, the 5'- and 3'-flanking sequences were determined to nucleotides -789 and 1602, respectively, and are available from GenBank.

Fragment 1 hybridized to an 834-bp band from an *EcoRI* digest of *B. subtilis* chromosomal DNA, and this fragment was isolated from a plasmid library by hybridization screening, yielding pPG13. Two rounds of plasmid rescue were required to complete the cloning of *guaA*. (i) To obtain the 3' end of the gene, fragment 2 in plasmid pPG13 was integrated into the *B. subtilis* chromosome and fragment 3 of about 2 kb was rescued in plasmid pGA13 after digestion with *PstI*. Copy number suppression in strain KE94 was required to subclone a 780-bp *EcoRI* fragment containing the 3' end of

-178 GAATTGATGAAAAATACGACATAGAACATGTCAAAGCTGACATTAAGCTATGAAAGGG
 -118 TCGCTTTACAGAAATGAAGATTCTTCGGCGCTATGGAATGATTTCGAGAGAGTTATGGGAA
 -35 -10 +1
 -58 AATTAAATTAAGAAGATGGTCTTGACCGCTTATCGACGTGTGTAGAAATTAGTGAAT
 3 ATTATCGGAGTCTGGGAGCGAGTTGGCCTGACTCGGCAACCGCCCTGCGCAAGAGGGC
 63 GGAGCGAGCTTCATATCTGCTCTGTTTTTTCTGTCTCAAAAATCTTAGAGGAGATTAG
 M T K L V N E M I L V L D F G S Q
 123 GTGCAACCATGACAAAGTTAGTGAATGAAATGATTCTGTCTGATTTCGCGAGTGCAG
 18 Y N Q L I T R R I R E F G V Y S E L H P
 183 TATAACAGCTGATTACACGCGCTATCCGTGAATTCGGTGTGTACAGCAGCTGCATCCA
 38 H T L T A E E I K K M N P K G I I L S Q
 243 CATACATTGACGCGTGAAGAAATTAAGAAATGAATCCAAAGGAATATTATTTATCCGGC
 58 G P N S V Y D E N S F R C D E K I F E L
 303 GGTCCAAACAGTGTGATGATGAAACTCTTTCGGCTGTGACGAGAAATCTTCGAGCTT
 78 D I P V L G I C Y G M Q L M T H Y L G G
 363 GATATCTCTGTTTGGGAATTTGCTACGCGTACGCTGATTGACTGATTCGCGCGGT
 98 K V E A A T Q R E Y G K A N I R I E G T
 423 AAAAGTTGAAGCGCAACGCGCGTGAATACGGAAGCAACATCCGCTATCGAAGGCA
 118 P D L F R D L P N E G V V W M S H G D L
 483 CCGATTCTGTTTTCAGAGATCTTCCGAATGAAGCAAGTGGTTGGATGAGCCGCGGATTG
 138 V V E V P E G F T V D A A T S H H C P N
 543 GTTGTAAGTCTCTGAAGGCTTCACTGTGACGCGCGCAAGCCATCACTGCCCGAAC
 158 S A M S K G D K K W H V Q F H P E V R
 603 TCAGCAATGACCAAGCGCAAAAAATGGCAGCGGTTCAGTTCACCCGCGAAGTGGCG
 178 H S E Y G N D L L K N F V F G V C E S E
 663 CACTCTGAATACGCAATGATCTTCTGAAAACTTGTATTCGGTGTTCGGAATCGGAA
 198 G E W S M E N F I E I E M I E R Q V
 723 GCGAATGGTCAATGGAAGCTTTATCGAAATCGAAATGCAAAAAATCCGTGAACAGGTC
 218 G D K Q V L C G L S G G V D S S V V A V
 783 GGAGACAAACAGGTTCTTTCGCGCTAAGCGCGCTGATGTTCTCTGTTGTTGCTGTT
 238 L I H K A I G D Q L T C G I F V H D G L L
 843 TTGATTCAAAACGATCGGCGACGAGTGTATCTTTGTAGACCATGCTCTCTC
 258 R K G G E R E G V M K T F S E G F N M N V
 903 CGTAAAGCGAAGCTGAGGTTGTTATGAAGAAATCAGCGAAGGCTTAAACATGAATGTG
 278 I K V D A K D R F L N K L K G V S D P E
 963 ATTAAAGTAGACGCAAAAGATCGATTCTTAAACAACTAAAGGCGTTTCTGATCTGAG
 298 Q K R K I I G N E F I Y V F D D E A D K
 1023 CAAAAACGCAAAATCATCGGTAAATGAAATTCATTGATGATGAAGCGGACAG
 318 L K G I D Y L A Q G T L Y T D I I E S G
 1083 CTCAAAGGATCGACTACCTTGCACAAGGTACGCTTTACACAGATATCATCGAGAGCGGT
 338 T A T A Q T I K S H H N V G G L P E D M
 1143 ACAGCAACGCGCAACGATCAATCTCACCACATGTGCGCGGATCTCTGTAAGACATG
 358 Q F E L I E P L N T L F K D E V R A L G
 1203 CAGTTTGAAGTGTGAGCGGTAAATACGCTCTTCAAGACGAAGTGCAGCGGCTTGGC
 378 T E L G I P D E I V W R Q P F P G P G L
 1263 ACAGAGCTCGGCATTCGGATGAAATCGTATGCGCTGAGCGGTTCACGAGGACCGGACTC
 398 G I R V L G E V T E E K L E I V R E S D
 1323 GGAATCCGCGTCTTGGCGAAGTAACAGAAAGAAATCTGAAATGTTCTGTAATCAGAT
 418 A I L R E E I A N H G L E R D I W Q V Y
 1383 GCAATCTCGGTGAAGAAATGCAATCAAGCTTAGAGCGTGATATCTGGCAATCTTCTC
 438 T V L P D I R S V G V M G D A R T Y D Y
 1443 ACTGTTCTCTGACATCCGACGCTGGTGTGTTGGGTGACGCAAGAACATATGATTAC
 458 T I E S R R T S I D G M T S D W A R I P
 1503 ACAATCGAATCGGCGGTACATCAATCGACGCGATGACATCTGACTGGGCGCGTATCCCG
 478 W D V L E V I S T R I V N E V K H I N R
 1563 TGGGATGCTTGAAGTCAATTCGACAGTATCGTAAACGAAGTGAAGCACATTAACCGC
 498 V V Y D I T S K P P A T I E W E
 1623 GTGGTGTATGATATTACAAGTAAGCGCGCTCGGACGATGAGTGGGAATAAGAATCAATT
 1683 AATGGAACCATCTTTTGGCAATTTGCGGGAAGATGGTTTATTTGTTTATTACGAA
 1743 CAAAATCGAATTTGCGGACTATTGTTGCTTTTGTATGTTTTCAAAATGAACCTT

FIG. 5. Nucleotide sequence of *guaA* and derived amino acid sequence of GMP synthetase. The nucleotide sequence is numbered from the start of transcription. Putative -35 and -10 promoter elements and a ribosome binding site (Shine-Dalgarno [SD]) are marked. Dyad symmetries with the potential for rho-independent transcription termination are marked in the 5' untranslated and 3'-flanking regions. A second dyad symmetry in the 5' untranslated region partially overlaps with the putative rho-independent terminator. Although not shown, the 5'- and 3'-flanking sequences were determined to nucleotides -359 and 1832, respectively, and are available from GenBank.

FIG. 6. Primer extension analysis of *purA* and *guaA* transcription start sites. Primers annealed to nucleotides 99 to 117 and 95 to 112 in *purA* and *guaA*, respectively. Dideoxy-sequencing ladders were obtained with the same primers used for cDNA synthesis. RNA was obtained from a prototrophic *B. subtilis* strain grown in minimal medium. Positions corresponding to the start of transcription are marked by an arrow: (left) *purA*, (right) *guaA*.

guaA from plasmid pPGA13 into pPGA16. (ii) For cloning of the 5' end of *guaA*, a 471-bp *Pvu*II segment of fragment 2 (plasmid pPG14) was integrated into chromosomal DNA and a 2.1-kb insert was rescued by cleavage with *Nco*I (fragment 4 in Fig. 3B). Southern blots of chromosomal DNA were consistent with a single *guaA* gene with restriction sites shown in Fig. 3B.

Nucleotide and derived amino acid sequences. The nucleotide sequences of *purA* and *guaA* in Fig. 4 and 5, respectively, were determined completely on both DNA strands. In *purA*, there is an open reading frame of 1,290 bp which encodes a protein of 430 amino acids. Gene *guaA* contains an open reading frame of 1,539 bp which could give a protein of 513 amino acids. There is a potential GTG translation initiation codon nine nucleotides upstream of the ATG shown in Fig. 5. The ATG was chosen based on the position of the putative Shine-Dalgarno sequence and comparison with *E. coli* *guaA* (see Discussion).

Mapping of promoters. Sites in *purA* and *guaA* for transcription initiation were mapped by primer extension. For each gene, two oligonucleotide primers were used to synthesize cDNA with reverse transcriptase. Figure 6 shows the results of primer extension mapping for *purA* and *guaA* with one set of primers. In each case, the same primer was used to generate the sequencing ladder. The start site for transcription of each gene is numbered +1 in Fig. 4 and 5. These transcription start sites were confirmed by using the second set of primers (data not shown). For each gene, the start of transcription is preceded by an appropriate σ^A promoter sequence. Dyad symmetries are noted in the 5' untranslated region of *guaA* and the 3'-flanking regions of *purA* and *guaA*.

DISCUSSION

The isolation of *purA* and *guaA* completes the cloning of genes for the 14 steps of de novo purine nucleotide synthesis in *B. subtilis*. Genes for the 10 steps from PRPP to IMP are clustered in a 12-gene *pur* operon, in which 11 of the genes encode enzymes for the biosynthetic pathway (3). Gene *purB* encodes adenylosuccinate lyase, which catalyzes step 8 in the pathway to IMP and also the final step in the branch

to AMP (Fig. 1). Gene *guaB*, encoding IMP dehydrogenase, the first enzyme for the branch to GMP, has been cloned (16). In contrast to *E. coli*, in which *guaB* (25) and *guaA* (26) are in an operon, these genes are at separate loci on the *B. subtilis* chromosome.

There are two potential σ^A -type promoters for *purA*: (–120) ATGACA-N₁₈-TATAAA (–91) and (–35) TTGACT-N₁₇-TAAACT (–7). Primer extension mapping of the transcription start site has provided evidence for function of the promoter sequence that is proximal to the operon. An mRNA transcribed with this promoter would have a 44-nucleotide 5' untranslated region preceding the site for translation initiation. Transcriptional fusions with a promoterless reporter gene are required to determine whether the more distal sequence has promoter function. The 5'-flanking region of *guaA* differs from that in *purA*. The *guaA* transcription start site was mapped to a position 130 nucleotides upstream of the codon for translation initiation. This site has an optimally positioned σ^A -dependent promoter, (–35) TTGACC-N₁₇-TAGAAT (–7).

Comparison of the 5'-flanking regions of *purA* and *guaA* with the *pur* operon suggests certain common regulatory features. The *pur* operon is subject to dual regulation of transcription by adenine and guanine compounds (3). Growth with excess adenosine results in repression of transcription initiation. Guanine compounds, on the other hand, regulate transcription by a termination-antitermination mechanism in a 242-nucleotide 5' untranslated region. ATP and guanine or hypoxanthine have been implicated as the coregulators for these steps (22). The 130-nucleotide *guaA* 5' untranslated region contains a potential secondary structure typical for factor-independent transcription termination which resembles that in the *pur* operon. This putative terminator could be part of a mechanism for guanine-mediated regulation of transcription. There is an additional potential overlapping secondary structure that could function in antitermination by precluding the formation of the terminator. In *purA*, the 44-nucleotide 5' untranslated leader mRNA does not contain sequences capable of secondary structure formation associated with termination-antitermination control. Repression of transcription initiation could potentially regulate *purA* expression. Addition of purines to cultures has been reported to repress levels of *purA*- and *guaB*-encoded enzymes, whereas the level of *guaA*-encoded GMP synthase remained constant (22). Based on correlations with nucleotide pools, it was suggested that the GTP/ATP pool ratio could regulate *guaB* and *purA* expression (22). Availability of cloned *purA*, *guaA*, and *guaB* should facilitate analyses of gene regulation.

Since *purA* and *guaA* were cloned independent of function, alignments with other sequences are important to confirm the identification. A sequence alignment of GMP synthetase from *E. coli* (26) and *Dictyostelium discoideum* (27) with the enzyme from *B. subtilis* is given in Fig. 7. The sequence from *D. discoideum* became available after completion of this work. In pairwise comparisons, 52% of the residues in GMP synthetase from *E. coli* and 41% of the aligned amino acids in the *D. discoideum* enzyme are identical to amino acids in the sequence from *B. subtilis*. These results establish the identification of *B. subtilis* GMP synthetase. The eukaryotic enzyme is about 200 amino acids larger than those from *B. subtilis* and *E. coli*, due primarily to an NH₂-terminal extension and several internal insertions. The NH₂-terminal 183 amino acids in *B. subtilis* GMP synthetase correspond to a G-type glutamine amide transfer domain that is conserved in six other amidotransferases, all

FIG. 8. Amino acid sequence alignment of *B. subtilis* (Bs), *E. coli* (Ec), mouse (M), and *D. discoideum* (Ds) adenylosuccinate synthetases. Highlighting in black is for residues that are conserved in at least three of the sequences.

of which are present in *B. subtilis* (3). Using degenerate primers to two conserved regions of the glutamine amide transfer domain, amino acid residues 84 to 90 and 177 to 183 in the *E. coli* sequence (Fig. 2 and 7), products amplified by PCR were recovered in three main bands after electrophoresis. Of 14 clones examined, four sequences were from *B. subtilis* GMP synthetase, two were 50% identical to *B. subtilis* GMP synthetase, and eight sequences were not recognized. We did not pursue the isolation of other amidotransferases from the complex mixture of products.

Sequences of adenylosuccinate synthetase from *E. coli* (30), *D. discoideum* (29), and the mouse (8) are aligned in Fig. 8 with the one from *B. subtilis* determined in this work. The identification of *B. subtilis* adenylosuccinate synthetase is confirmed by the identities of greater than 35% with the other sequences. Although a *D. discoideum* cDNA encoding GMP synthetase was isolated by functional complementation of an *E. coli guaB* mutant (27), multicopies of the intact *B. subtilis* gene appear to be lethal in *E. coli*. Because of lethality in *E. coli*, we have not succeeded in constructing

intact *guaA* and *purA* clones and obtaining translational fusions to *lacZ* to examine gene expression.

ACKNOWLEDGMENTS

We thank C. M. Thomas for supplying the *B. subtilis* gene library in a low-copy-number vector and Gaochao Zhou for help with methodology and computer analyses.

This work was supported by Public Health Service grant GM24658 from the National Institutes of Health. Computer facilities were supported by NIH grant AI 27713.

REFERENCES

1. Anagnostopoulos, C., and J. Spizizen. 1961. Requirements for transformation in *Bacillus subtilis*. J. Bacteriol. 81:741-746.
2. Birnboim, H. C., and J. Doly. 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. 7:1513-1523.
3. Ebbole, D. J., and H. Zalkin. 1987. Cloning and characterization of a 12-gene cluster from *Bacillus subtilis* encoding nine enzymes for *de novo* purine nucleotide synthesis. J. Biol. Chem. 262:8274-8287.
4. Ebbole, D. J., and H. Zalkin. 1988. Detection of *pur* operon-attenuated mRNA and accumulated degradation intermediates in *Bacillus subtilis*. J. Biol. Chem. 263:10894-10902.
5. Ebbole, D. J., and H. Zalkin. 1989. *Bacillus subtilis pur* operon expression and regulation. J. Bacteriol. 171:2136-2141.
6. Ebbole, D. J., and H. Zalkin. 1989. Interaction of a putative repressor protein with an extended control region of the *Bacillus subtilis pur* operon. J. Biol. Chem. 264:3553-3561.
7. Feinberg, A. P., and B. Vogelstein. 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. 132:6-13.
8. Guicherit, O. M., F. B. Rudolph, R. E. Kellems, and B. F. Cooper. 1991. Molecular cloning and expression of a mouse muscle cDNA encoding adenylosuccinate synthetase. J. Biol. Chem. 266:22582-22587.
9. Hanahan, D. 1983. Studies on transformation of *Escherichia coli* with plasmids. J. Mol. Biol. 166:557-580.
10. Hasnian, S., and C. M. Thomas. 1986. Construction of a novel gene bank of *Bacillus subtilis* using a low copy number vector in *Escherichia coli*. J. Gen. Microbiol. 132:1863-1874.
11. Kanzaki, N., and K. Miyagawa. 1990. Nucleotide sequence of the *Bacillus subtilis* IMP dehydrogenase gene. Nucleic Acids Res. 18:6710.
12. Lopilato, J., S. Bortner, and J. Beckwith. 1986. Mutation in a new chromosomal gene of *Escherichia coli* K-12, *pcnB*, reduces plasmid copy number of pBR322 and its derivatives. Mol. Gen. Genet. 205:285-290.
13. Mandel, M., and A. Higa. 1970. Calcium-dependent bacteriophage DNA infection. J. Mol. Biol. 53:159-162.
14. Messing, J. 1983. New M13 vectors for cloning. Methods Enzymol. 101:20-78.
15. Miller, J. H. 1972. Experiments in molecular genetics, p. 433. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
16. Miyagawa, K., H. Kimura, K. Nakahama, M. Kibuchi, M. Doi, S. Akiyama, and Y. Nakao. 1986. Cloning of the *Bacillus subtilis* IMP dehydrogenase gene and its application to increased production of guanosine. Bio/Technology 4:225-228.
17. Momose, H., H. Nishikawa, and I. Shio. 1966. Regulation of purine nucleotide synthesis in *Bacillus subtilis*. 1. Enzyme repression by purine derivatives. J. Biochem. 59:325-331.
18. Nishikawa, H., H. Momose, and I. Shio. 1967. Regulation of purine nucleotide synthesis in *Bacillus subtilis*. 2. Specificity of purine derivatives for enzyme repression. J. Biochem. 62:92-98.
19. Saito, H., and K. I. Miura. 1963. Preparation of transforming deoxyribonucleic acid by phenol treatment. Biochim. Biophys. Acta 72:619-629.
20. Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. Molecular cloning: a laboratory manual, 2nd ed., p. 630. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
21. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74:5463-5467.
22. Saxild, H. H., and P. Nygaard. 1991. Regulation of levels of purine biosynthetic enzymes in *Bacillus subtilis*: effects of changing purine nucleotide pools. J. Gen. Microbiol. 137:2387-2394.
23. Schwartz, R. M., and M. O. Dayhoff. 1979. Dayhoff table, p. 353-358. In M. O. Dayhoff (ed.), Atlas of protein sequence and structure. National Biomedical Research Foundation, Washington, D.C.
24. Tabor, S., and C. C. Richardson. 1987. DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. Proc. Natl. Acad. Sci. USA 84:4767-4771.
25. Tiedeman, A. A., and J. M. Smith. 1985. Nucleotide sequence of *guaB* locus encoding IMP dehydrogenase of *Escherichia coli* K12. Nucleic Acids Res. 13:1303-1316.
26. Tiedeman, A. A., J. M. Smith, and H. Zalkin. 1985. Nucleotide sequence of the *guaA* gene encoding GMP synthetase of *Escherichia coli* K12. J. Biol. Chem. 260:8676-8679.
27. Van Lookeren Campagne, M. M., J. Franke, and R. H. Kessin. 1991. Functional cloning of *Dictyostelium discoideum* cDNA encoding GMP synthetase. J. Biol. Chem. 266:16448-16452.
28. Wang, R. F., and S. R. Kushner. 1991. Construction of versatile low-copy-number vectors for cloning, sequencing and gene expression in *Escherichia coli*. Gene 100:195-199.
29. Weismuller, L., J. Wittbrodt, A. A. Noegel, and M. Schleicher. 1991. Purification and cDNA-derived sequence of adenylosuccinate synthetase from *Dictyostelium discoideum*. J. Biol. Chem. 266:2480-2485.
30. Wolf, S. A., and J. M. Smith. 1988. Nucleotide sequence and analysis of the *purA* gene encoding adenylosuccinate synthetase of *Escherichia coli* K12. J. Biol. Chem. 263:19147-19153.
31. Zhou, G., J. E. Dixon, and H. Zalkin. 1990. Cloning and expression of avian glutamine phosphoribosylpyrophosphate amidotransferase. Conservation of a bacterial propeptide sequence supports a role for posttranslational processing. J. Biol. Chem. 265:21152-21159.