Nucleotide sequence of the *guaB* locus encoding IMP dehydrogenase of *Escherichia coli* K12

Amelia A.Tiedeman and John M.Smith*

Department of Biochemistry and Molecular Biology, Louisiana State University Medical Center,
Shreveport, LA 71130, USA

Received 14 January 1985; Accepted 23 January 1985

ABSTRACT

IMP dehydrogenase, the product of the *guaB* locus in *Escherichia coli* K12, catalyzes the synthesis of XMP by the NAD⁺ dependent oxidation of IMP. The *guaB* locus has been subcloned from the Clarke and Carbon plasmid pLC34-10. The sequence of the *guaB* structural gene and surrounding DNA was determined by the dideoxy chain termination method of Sanger. The 1.533 kb *guaB* gene encodes an IMP dehydrogenase subunit of molecular weight 54,512. *St* nuclease mapping placed the site of *guaBA* mRNA initiation approximately 188 bp from the start of the *guaB* structural gene. The -10 and -35 regions that define the *guaBA* promoter were located upstream of the start of the *guaBA* transcription initiation site. The control region of approximately 188 bp does not show any obvious potential for secondary structure. A secondary λ att site has been identified 42 bp distal to the *guaB* start codon.

INTRODUCTION

IMP dehydrogenase (EC 1.2.1.14) catalyzes the NAD⁺ dependent conversion of IMP to XMP (1), the first reaction unique to GMP biosynthesis. The enzyme from *Escherichia coli* has been purified and characterized as a tetramer of identical subunits (2,3). IMP dehydrogenase from other bacterial sources has been shown to vary widely with respect to allosteric properties, size and subunit composition (4,5,6,7,8,9,10,11). The molecular weight of the *E. coli* IMP dehydrogenase subunit has been reported as 54,000 (3) and 58,000 (2); it contains a cysteine at the IMP binding site (4) and is inhibited in a simple competitive manner by GMP (2,3). The *E. coli* IMP dehydrogenase does not exhibit allosteric properties as does the IMP dehydrogenase from *Bacillus subtilis* (10) or *Salmonella typhimurium* (11).

In *E. coli* and *S. typhimurium*, IMP dehydrogenase is the product of the *guaB* gene, the operator proximal gene in the *guaBA* operon (12,13). The *guaB* and *guaA* genes are coordinately regulated (12,14); however, a secondary promoter for the *guaA* gene has been localized to the end of the *guaB* gene (15). The *guaBA* operon is subject to dual control by adenine and guanine nucleotides; it is repressed by guanine nucleotides and induced by adenine nucleotides (16,17,18).

The only regulatory mutations reported to affect the expression of the guaBA operon have been isolated through the use of gua-lac fusion strains (19,20). These regulatory mutations have been divided into three classes: linked cis-acting, linked trans-acting (guaR) and unlinked trans-acting regulatory mutations (20).

Since it will be necessary to know the DNA sequence of the control region to fully understand the dual regulatory mechanism of the guaBA operon, we have subcloned and sequenced the guaB structural gene and its surrounding DNA. We have determined that the coding region of the guaB gene spans 1.533 kb and codes for an IMP dehydrogenase subunit of molecular weight 54,512. S1 nuclease mapping placed the site of guaBA mRNA initiation approximately 188 bp upstream of the start of the guaB structural gene. We have also identified the -10 and -35 regions that define the guaBA promoter.

MATERIALS AND METHODS

Strains and Media

The $\Delta(\text{xse-guaB})91$ (13) and recA56 (20) mutations were moved into one of our laboratory strains by standard genetic techniques (21) to create strain TX506 [$\Delta(\text{xse-guaB})91$ ara $\Delta(\text{lac})$ guaC srlC300::Tn10 recA56]. This strain was used in the isolation of guaB plasmids by complementation. The wild type strain W3110 (22) was employed as a source of in vivo guaBA mRNA. Strain JM83 (23) was used as a recipient to identify subclones containing restriction fragments from the guaB region while strain JM101 (23) was employed for the propagation of M13 bacteriophages. The minimal medium of Neidhardt (24) and the rich media described by Miller (21) were used for the growth of the E. coli K12 strains. Strains were made competent and transformed by the procedure of Dagert and Ehrlich (25).

Plasmids

Plasmid pLC34-10 from the Clarke and Carbon library (26), previously identified as containing the guaBA operon (27), was obtained from the Coli Genetic Stock Center, Yale University, New Haven, CT. Plasmids pUC9 (28) and pUC12 (29) were employed for the subcloning of restriction fragments. Restriction fragment sizes were determined by the program of Schaffer (30) that was translated into Pascal and executed on an Apple II Plus computer.

DNA Isolation

The procedure of Holmes and Quigly (31) was employed for rapid plasmid isolations from 5 ml E. coli K12 cultures. The NACS-37 form of the RPC-5 analog resin (32) (Bethesda Research Labs, Gaithersburg, MD) was used for

the preparative scale isolation of plasmid DNA. The manufacturers recommended protocols were employed with the following modifications: RNase A was used instead of RNase T1, a Sephacryl S-300 chromatography step (33) was added to remove digested RNA from the plasmid preparation before application to the RPC-5 Analog column and the plasmid DNA was eluted by a step instead of a linear gradient. DNA fragments were recovered from low melting temperature agarose gels as described by Guo *et al.* (34).

DNA Sequence Analysis

DNA sequences were determined by the dideoxy chain termination method of Sanger *et al.* (35). The complete DNA sequence of a 2.029 kb DNA fragment containing the intact *guaB* gene was determined for both strands from overlapping DNA sequences. A detailed restriction map of the *guaB* locus and sequencing strategy are shown in Fig. 3. The indicated restriction fragments were ligated into the appropriate cloning sites in M13mp8 and M13mp9 (36) or M13mp10 and M13mp11 (29) and transformed into JM101 (29). Colorless plaques were individually picked and propagated for the preparation of DNA (37). The 17 base primer (5'GTTTCCAGTCACGAC-3') used for sequencing was kindly supplied by B.A. Roe, University of Oklahoma. The DNA sequences were compiled and analyzed by computer (38).

RNA Studies

RNA was extracted from strain W3110 by the SDS/hot phenol method (39). The S1 nuclease procedure of Berk and Sharp (40) was used to determine the 5' end of the *guaB* mRNA. A 109 bp TaqI fragment that spanned the start of the *guaB* gene and a 188 bp EcoRI-TaqI fragment were 5' end labeled with [γ - 32 P]-ATP by polynucleotide kinase after dephosphorylation with calf alkaline phosphatase as described by Maniatis *et al.* (41). 150 μ g of RNA from strain W3110 was hybridized to 100 ng of each labeled fragment for 18 hours at 51°C. After division of the hybridization mixtures into 3 equal portions, 100, 300, and 700 units of S1 nuclease were added and the mixtures incubated for 30 min at 37°C. The reactions were terminated by phenol extraction, ethanol precipitated, resuspended in loading buffer, heated at 100°C for 7 min and then immediately loaded onto a sequencing gel. The size of the protected fragments was determined after autoradiography by comparison to an accompanying Sanger dideoxy sequencing ladder and adjusted by 0.5 nucleotide (42).

Enzymes and Chemicals

[α - 32 P]-dATP was obtained from New England Nuclear (Boston, MA) and [γ - 32 P]-ATP was obtained from ICN (Irvine, CA). T4 DNA ligase, DNA

Figure 1. Partial Restriction Map and Complementation Analysis of Plasmid pLC34-10. The thick line represents chromosomal DNA and the thin line represents ColE1 DNA. Only the restriction sites pertinent to the *guaBA* operon and this study are shown. Complementation studies were carried out in strains TX506 [$\Delta(\underline{xse-guaB})$] and TX538 (*guaA*). N.T., Not Tested.

polymerase I (Klenow fragment) and SmaI restriction enzyme were obtained from Boehringer Mannheim (Indianapolis, IN). SI nuclease was obtained from BRL (Gaithersburg, MD). All other enzymes were obtained from New England Biolabs (Beverly, MA). Deoxy- and dideoxy nucleotide triphosphates were obtained from Pharmacia PL Biochemicals (Milwaukee, WI). 5-Bromo-4-chloro-3-indoyl- β -D-galactoside and all other chemicals were obtained from Sigma (St. Louis, MO). Sea Plaque low melting temperature agarose was obtained from FMC Corp (Rockland, ME).

RESULTS

Cloning of *guaB*

Plasmid pLC34-10 from the Clarke and Carbon library (26) has been previously identified as containing the *guaBA* operon (27). A preliminary restriction map of pLC34-10 was determined for the restriction enzymes PstI, EcoRI and SmaI (Fig. 1). Plasmid pLC34-10 was calculated to be approximately 15 kb which reflects a chromosomal insert of 8.6 kb. DNA fragments generated by the PstI or XmaI (isoschizomer of SmaI) restriction enzymes were subcloned into pUC9 and transformed into strain TX506 [$\Delta(\underline{xse-guaB})$]. Only subclones containing either orientation of the 8.1 kb XmaI fragment were capable of *guaB* complementation. Restriction analysis of a representative plasmid containing the 8.1 kb XmaI fragment showed the presence of the unique EcoRI site of pLC34-10. A comparison of the restriction map generated by Eisenbeis and Parker (43) for the neighboring *hisS* locus and our restriction map of pLC34-10 revealed several common restriction sites and

Figure 2. Bal31 Nuclease Shortening of the 6.0 kb ClaI-SmaI Fragment. Plasmid pJS32 was linearized with HindIII and then digested with Bal31 (34) to remove approximately 3.0 kb of DNA (dashed line). After repair of the DNA ends with DNA polymerase I (Klenow fragment) (41), the shortened DNA was digested with SalI and electrophoresed on a 0.7% low melting temperature agarose gel. 2.9 kb fragments were recovered and ligated into SmaI-SalI sites of plasmid pUC12. *Gua*⁺ transformants were selected in strain TX506 [$\Delta(\text{xse-guaB})$].

fragment sizes. Thus, we were able to orient the inserted chromosomal DNA of pLC34-10 with respect to the *guaBA* operon and the *E. coli* chromosome. Since further restriction analysis of the 8.1 kb XmaI fragment failed to locate any additional unique restriction sites useful for subcloning, we decided to employ a Bal31 nuclease treatment to shorten the *guaB* DNA insert.

In preparation, we repeated the subcloning of *guaB* from pLC34-10 by using a ClaI site present in the ColEI portion of pLC34-10 and the SmaI site within the *guaA* structural gene (Fig. 1). The 6.0 kb ClaI-SmaI fragment was ligated into the AccI-SmaI sites of pUC9 to form plasmid pJS32 (Fig. 2). We have sequenced *guaA* (Tiedeman *et al.*, submitted for publication) and located the SmaI site 1.2 kb distal to the start of the *guaA* gene. Estimating

that an IMP dehydrogenase subunit of MW 54,000 would require a minimum of 1.7 kb of coding DNA, we calculated that the 6.0 kb ClaI-SmaI fragment contained approximately 3.0 kb of DNA beyond the start of the guaB gene. Thus, Bal31 nuclease was used to remove approximately 2.8 kb of DNA from the ClaI end of the 6.0 kb ClaI-SmaI fragment as shown in Fig. 2. Small scale plasmid harvests from guaB⁺ transformants of strain TX506 [$\Delta(\text{xse-guaB})$] were characterized by EcoRI digestion to determine the size of the shortened fragment. The smallest EcoRI fragment found was approximately 800 bp and should reflect the minimum amount of DNA beyond the guaB structural gene necessary for expression. A plasmid with a slightly larger insert was chosen for sequencing and designated pJS49.

Nucleotide Sequence of the DNA Fragment Encoding guaB

The nucleotide sequence of the guaB gene and the surrounding DNA is shown in Fig. 4. The Bal31 digestion stop point is designated as the 5' end. The data in Fig. 4 show 240 bp of sequence upstream of the guaB gene, 1,533 bp encoding the guaB structural gene and 68 bp of intercistronic region to the start of the guaA structural gene.

Derived Amino Acid Sequence of IMP Dehydrogenase

The DNA sequence shown in Fig. 4 contains a single open reading frame of 1,533 nucleotides starting with an ATG at nucleotide position 242. This ATG initiation codon is preceeded by a good Shine-Dalgarno sequence (44) that obeys Rules 4 and 5 of Stormo *et al.* (45). The next several ATG codons encountered in the sequence are not preceeded by acceptable Shine-Dalgarno regions. The primary translation product contains 511 amino acids and has a calculated molecular weight of 54,512. The amino acid sequence of IMP dehydrogenase has not been determined so a direct confirmation of the nucleotide sequence is not possible.

However, the following lines of evidence support this proposed sequence. The molecular weight of the protein encoded by the open reading frame is 54,512. This is in close agreement with the two published values of 54,000 (3) and 58,000 (2) for the purified IMP dehydrogenase subunit size. An amino acid analysis of IMP dehydrogenase has been reported (3). Our DNA sequence derived amino acid composition of IMP dehydrogenase is shown in Table 1. Gilbert *et al.* (2) reported a total of 55 lysine and arginine residues for purified IMP dehydrogenase which is in excellent agreement with our DNA sequence derived total of 57. The N-terminal amino acid of IMP dehydrogenase has been reported as methionine (2) and thus would indicate that the initiator methionine has not been processed. Thomas and Drabble (46)

Table 1. Amino Acid Composition and Codon Usage of IMP Dehydrogenase

Codon Usage			
Phe TTT 5	Ser TCT 7	Tyr TAT 5	Cys TGT 5
Phe TTC 4	Ser TCC 12	Tyr TAC 5	Cys TGC 1
Leu TTA 2	Ser TCA 1	End TAA 0	End TGA 1
Leu TTG 1	Ser TCG 1	End TAG 0	Trp TGG 0
Leu CTT 1	Pro CCT 3	His CAT 0	Arg CGT 22
Leu CTC 5	Pro CCC 1	His CAC 10	Arg CGC 10
Leu CTA 1	Pro CCA 1	Gln CAA 5	Arg CGA 0
Leu CTG 32	Pro CCG 13	Gln CAG 12	Arg CGG 1
Ile ATT 12	Thr ACT 11	Asn AAT 2	Ser AGT 1
Ile ATC 20	Thr ACC 16	Asn AAC 9	Ser AGC 6
Ile ATA 2	Thr ACA 2	Lys AAA 24	Arg AGC 0
Met ATG 15	Thr ACG 7	Lys AAG 0	Arg AGG 0
Val GTT 23	Ala GCT 18	Asp GAT 6	Gly GGT 28
Val GTC 5	Ala GCC 8	Asp GAC 17	Gly GGC 24
Val GTA 6	Ala GCA 13	Glu GAA 33	Gly GGA 0
Val GTG 16	Ala GCG 13	Glu GAG 8	Gly GGG 0
Total Codons Used = 511			
Amino Acid Composition			
Ala 52	Gln 17	Leu 42	Ser 28
Arg 33	Glu 41	Lys 24	Thr 36
Asn 11	Gly 52	Met 15	Trp 0
Asp 23	His 10	Phe 9	Tyr 10
Cys 6	Ile 34	Pro 18	Val 50
Molecular Weight = 54,512			

have reported a molecular weight of 26,000 for a polypeptide initiating at the *guaB* promoter and ending at the internal *EcoRI* site shown in Fig. 3. In close agreement, our DNA sequence predicts a value of 23,514 for this fragment. Trypsin digestion of IMP dehydrogenase yields polypeptides of molecular weight 42,000 and 10,000 (2). Within our deduced amino acid sequence, cleavage at any of the three lysine residues (lys-372,-426,-439) would yield polypeptides of the approximate size reported. Finally, the *PvuII* site at position 407 is predicted to cleave in phase of the *guaB* reading frame. This was confirmed by cloning the *EcoRI*-*PvuII* fragment into the *EcoRI*-*SmaI* sites of the promoter-translation detection plasmid pMC1403 (47) to produce an active β -galactosidase. A *Bal31* shortened DNA fragment cloned into the *EcoRI*-*SmaI* site of pMC1403 and

Figure 3. Sequencing Strategy for the *guaB* Gene. The arrows denote sequencing direction and length of sequences obtained in individual experiments. Bal31 designates the end of a Bal31 nuclease shortened fragment. The first EcoRI site, originally from pUC12, is designated as the 5' end.

selected for β -galactosidase expression was also found to be in the predicted reading frame. Thus, we conclude that the *guaB* DNA sequence encoding IMP dehydrogenase is correct.

Codon Usage

The codon usage of the *guaB* structural gene is reported in Table 1. As expected, the codon usage in *guaB* matches that of a highly expressed gene (48,49). However, there are four features worth noting; only the AAA codon is used to code for all 24 lysine residues; CAC is the sole codon used for the 10 histidine residues; the cysteine codon TGT is more highly used than TGC and the DNA sequence predicts that IMP dehydrogenase contains no tryptophan residues.

Location of the Secondary λ att Site

Secondary lambda site insertions into the *guaB* gene have been reported to be at the same site and have been mapped to the promoter proximal end of the *guaB* gene (50). Recently, Thomas and Drabbe (46) have further localized the secondary λ attachment site 40 to 50 bp from the start of the *guaB* gene. A search of our DNA sequence reveals only one good candidate for a secondary λ attachment site in this region starting at position 278. The 15 base

wild type attB core sequence (59) and the proposed guaB att are aligned for maximum homology and to preserve the block of 3 T residues at -2, -1 and 0 common to secondary λ att sites (60).

	-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7
<u>attB</u>	G	C	T	T	T	T	T	A	T	A	C	T	A	A	
<u>guaB</u>	C	A	A	T	A	T	T	T	A	T	T	A	A	C	C

The secondary site located in guaB is responsible for approximately 0.025% of the lysogens isolated from attB deletion strains (50).

S1 Nuclease Studies

The 5' end of the *in vivo* guaBA mRNA was determined by S1 nuclease mapping. A 109 bp TaqI fragment that spanned the start of the guaB structural gene and a 188 bp EcoRI-TaqI fragment (Fig. 3) were initially chosen for S1 nuclease mapping. The fragments were dephosphorylated with calf alkaline phosphatase and end-labeled with [γ - 32 P]-ATP and polynucleotide kinase. The labeled fragments were then denatured and hybridized to *in vivo* mRNA from the wild type strain W3110. After S1 nuclease digestion, the protected fragments were sized on a DNA sequencing gel with a Sanger dideoxy sequencing ladder as a standard. The 109 bp TaqI fragment was completely protected and yielded a single band. The 188 bp EcoRI-TaqI fragment yielded one major and 3 minor protected bands of 130 to 134 nucleotides (data not shown). This places the start of transcription of the guaBA operon at position 54 to 57 of the sequence shown in Fig. 4. It is not known if the multiple bands represent multiple initiations or are degradation products of the *in vivo* guaBA mRNA.

Identification of the guaBA Promoter

The DNA sequence upstream of nucleotides 54 to 57 was searched for potential -10 and -35 regions. The best candidates for the guaBA -10 and -35 regions are overlined in Fig. 5. The proposed -10 region (TAGAAA) differs from the consensus sequence TATAAT (51) at two positions and the -35 region differs at two positions (GTGACC versus consensus TTGACA). Another potential -10 region partially overlaps the underlined -10 region but does not maintain the normal spacing with the guaBA mRNA initiation region or an appropriate -35 region.

DNA Sequence Flanking the guaB Gene

The 241 nucleotides of DNA sequence preceeding the guaB structural gene and the 68 bp guaB-guaA intercistronic region are shown in Fig. 5. While the region preceeding the guaB structural gene does not contain any obvious secondary structure, it is interesting to note that the -10 sequence is centered in a region of dyad symmetry (GAAATTAGAAATTC) that also contains a

2
TAAAGTACCAAGTACCGGAAGCTGGTTGCGTGAAATTAGAAATTTGCCCGTGATCCAAACCTGTCCCATCTCATGCTCA

82
AGCAGCAGACGAACCGTTTGATTAGCGGACTAACGGTAAAAATTGCAGGGGATTGAGAAGTAACATGTGAGCGAGATC

162
AAATTCTAAATCAGCAGGTATTTCAGTCGATAGTAACCCGCCCTTCGGGGATAGCAAGCATTTTTGCAAAAAGGGGTAG

242
MET GLN SER VAL THR LEU CYS ILE MET PRO ARG GLN TYR LEU LEU THR THR LEU VAL GLU
ATG CAA TCG GTT ACG CTC TGT ATA ATG CCG CGG CAA TAT TTA TTA ACC ACT CTG GTC GAG

302
ILE LEU PRO MET LEU ARG ILE ALA LYS GLU ALA LEU THR PHE ASP ASP VAL LEU LEU VAL
ATA TTG CCC ATG CTA CGT ATC GCT AAA GAA GCT CTG ACG TTT GAC GAC GTT CTC CTC GTT

362
PRO ALA HIS SER THR VAL LEU PRO ASN THR ALA ASP LEU SER THR GLN LEU THR LYS THR
CCT GCT CAC TCT GAT GTT CTG CCG AAT ACT GCT GAC CTC AGC ACC CAG CTG ACG AAA ACT

422
ILE ARG LEU ASN ILE PRO MET LEU SER ALA ALA MET ASP THR VAL THR GLU ALA ARG LEU
ATT CGT CTG AAT ATC CCT ATG CTT TCC GCA GCA ATG GAT ACC GTA ACG GAA GCG GCG CTG

482
ALA ILE ALA LEU ALA GLN GLU GLY GLY ILE GLY PHE ILE HIS LYS ASN MET SER ILE GLU
GCT ATT GCT CTG GCT CAG GAA GGC GGT ATC GGC TTT ATC CAC AAA AAC ATG TCC ATT GAA

542
ARG GLN ALA GLU GLU VAL ARG ARG VAL LYS LYS HIS GLU SER GLY VAL VAL THR ASP PRO
CGC CAG GCA GAA GAA GTT CGC CGT GTG AAA AAA CAC GAA TCT GGT GTG GTG ACT GAT CCG

602
GLN THR VAL LEU PRO THR THR THR LEU ARG GLU VAL LYS GLU LEU THR GLU ARG ASN GLY
CAG ACT GTT CTG CCA ACC ACG ACG CTG CGC GAA GTG AAA GAA CTG ACC GAG CGT AAC GGT

662
PHE ALA GLY TYR PRO VAL VAL THR GLU GLU ASN GLU LEU VAL GLY ILE ILE THR GLY ARG
TTT GCG GGC TAT CCG GTC GTT ACC GAA GAA AAC GAA CTG GTG GGT ATT ATC ACC GGT CGT

722
ASP VAL ARG PHE VAL THR ASP LEU ASN GLN PRO VAL SER VAL TYR MET THR PRO LYS GLU
GAC GTG CGT TTT GTT ACC GAC CTG AAC CAG CCG GTT AGC GTT TAC ATG ACG CCG AAA GAG

782
ARG LEU VAL THR VAL ARG GLU GLY GLU ALA ARG GLU VAL VAL LEU ALA LYS MET HIS GLU
CGT CTG GTC ACC GTG CGT GAA GGT GAA GCC CGT GAA GTG CTG CTG GCA AAA ATG CAC GAA

842
LYS ARG VAL GLU LYS ALA LEU VAL VAL ASP ASP GLU PHE HIS LEU ILE GLY MET ILE THR
AAA CCG GTT GAA AAA GCG CTG GTG GTT GAT GAC GAA TTC CAC CTG ATC GGC ATG ATC ACC

902
VAL LYS ASP PHE GLN LYS ALA GLU ALA LYS PRO ASN ALA CYS LYS ASP GLU GLN GLY ARG
GTG AAA GAC TTC CAG AAA GCG GAA GCT AAA CCG AAC GCC TGT AAA GAC GAG CAA GGC CGT

962
LEU ARG VAL GLY ALA ALA VAL GLY ALA GLY ALA GLY ASN GLU GLU ARG VAL ASP ALA LEU
CTG CGT GTT GGT GCA GCG GTT GGC GCA GGT GCG GTT AAC GAA GAG CGT GTT GAC GCG CTG

1022
VAL ALA ALA GLY VAL ASP VAL LEU LEU ILE ASP SER SER HIS GLY HIS SER GLU GLY VAL
GTT GCC GCA GGC GTT GAC GTT CTG CTG ATC GAC TCC TCC CAC GGT CAC TCA GAA GGT GTA

1082
LEU GLN ARG ILE ARG GLU THR ARG ALA LYS TYR PRO ASP LEU GLN ILE ILE GLY GLY ASN
CTG CAA CGT ATC CGT GAA ACC CGT GCT AAA TAT CCG GAT CTG CAA ATT ATC GGC GGC AAC

1142
VAL ALA THR ALA ALA GLY ALA ARG ALA LEU ALA GLU ALA GLY CYS SER ALA VAL LYS VAL
GTG GCA ACA GCT GCA GGT GCA GCG GCT CTG GCA GAA GCT GGT TGC AGT GCG GTT AAA GTC

1202
GLY ILE GLY PRO GLY SER ILE CYS THR THR ARG ILE VAL THR GLY VAL GLY VAL PRO GLN
GGC ATT GGC CCT GGC TCT ATC TGT ACA ACT CGT ATC GTG ACT GGC GTC GGT GTT CCG CAG

1262
ILE THR ALA VAL ALA ASP ALA VAL GLU ALA LEU GLU GLY THR GLY ILE PRO VAL ILE ALA
ATT ACC GCT GTT GCT GAC GCA GTA GAA GCC CTG GAA GGC ACC GGT ATT CCG GTT ATC GCT

1322
ASP GLY GLY ILE ARG PHE SER GLY ASP ILE ALA LYS ALA ILE ALA ALA GLY ALA SER ALA
GAT GGC GGT ATT GCG TTC TCC GGC GAC ATC GCC AAA GCT ATC GCC GCT GGC GCA AGC GCG

1382
VAL MET VAL GLY SER MET LEU ALA GLY THR GLU GLU SER PRO GLY GLU ILE GLU LEU TYR
GTG ATG GTA GGT TCC ATG CTG GCG GGT ACT GAA GAA TCT CCG GGT GAA ATC GAA CTC TAC

1442
GLN GLY ARG SER TYR LYS SER TYR ARG GLY MET GLY SER LEU GLY ALA MET SER LYS GLY
CAG GGC CGT TCT TAC AAA TCT TAC CGT GGT ATG GGT TCC CTG GGC GCG ATG TCC AAA GGT

1502
SER SER ASP ARG TYR PHE GLN SER ASP ASN ALA ALA ASP LYS LEU VAL PRO GLU GLY ILE
TCC TCT GAC CGT TAT TTC CAG AGC GAT AAC GCT GCC GAC AAA CTG GTG CCG GAA GGT ATC

```

1562
GLU GLY ARG VAL ALA TYR LYS GLY ARG LEU LYS GLU ILE ILE HIS GLN GLN MET GLY GLY
GAA GGT CGC GTA GCC TAT AAA GGT CGC CTG AAA GAG ATC ATT CAC CAG CAG ATG GGC GGC

1622
LEU ARG SER CYS MET GLY LEU THR GLY CYS GLY THR ILE ASP GLU LEU ARG THR LYS ALA
CTG CGC TCC TGT ATG GGT CTG ACC GGC TGT GGT ACT ATC GAC GAA CTG CGT ACT AAA GCG

1682
GLU PHE VAL ARG ILE SER GLY ALA GLY ILE GLN GLU SER HIS VAL HIS ASP VAL THR ILE
GAG TTT GTA CGT ATC AGC GGT GCG GGC ATT CAG GAA AGC CAC GTT CAC GAC GTG ACC ATT

1742
THR LYS GLU SER PRO ASN TYR ARG LEU GLY SER TTCTCTTCGCCGACTTCATGTCGGGCGATT
ACT AAA GAG TCC CCG AAC TAC CGT CTG GGC TCC TGA

1809
TATATTATCTGTTTCACTTGCCCTCGGAATAAGCGTCAATG

```

Figure 4. DNA Sequence and Translation of *guaB*. The sequence of the sense strand of the *guaB* gene is shown. The sequence is numbered from the *Bal31* digestion stop point as the 5' end and includes 241 nucleotides upstream of *guaB*, 1,533 nucleotides of coding region as well as the 68 nucleotide intercistronic region. The terminal ATG is the initiator codon of the *guaA* gene (Tiedeman *et al.*, submitted for publication). The proposed Shine-Delgarno sequence (44) is boxed and the secondary λ att site is underlined.

Figure 5. DNA Sequence of the Non-Coding Regions Surrounding the *guaB* Structural Gene. The -10 and -35 regions as well as the start of the *guaB* transcript are overlined. The RNA transcript of the *guaB*-*guaA* intercistronic region has the potential to form a hairpin loop with a calculated ΔG of -23 kcal (61). The two underlined sequences in the intercistronic region are potential -10 regions for the *guaA* secondary promoter (Tiedeman, *et al.*, submitted for publication).

direct repeat (GAAATT) (Fig. 5). One other region of notable dyad symmetry (TTTTJGCAAAAA) immediately precedes the guaB initiator codon. The guaB structural gene ends with a 68 bp intercistronic region that has the potential to form an RNA structure typical of a rho-independent terminator with a ΔG of -23 Kcal.

DISCUSSION

The DNA sequence of the guaB gene was determined from a restriction fragment subcloned from plasmid pLC34-10. The guaB gene was identified by its location adjacent to guaA, the molecular weight of its product, and by in phase guaB-lacZ gene fusions. The DNA sequence predicts a molecular weight of 54,512 for the IMP dehydrogenase subunit and a molecular weight of 218,048 for the tetrameric holoenzyme. The secondary λ att site in guaB was located approximately 40 bp distal to the guaB start codon.

The start of the guaB mRNA was determined by S1 nuclease mapping to lie approximately 188 bp upstream of the guaB initiator codon. It is unclear from our data whether the heterogeneity at the 5' end of the guaB mRNA reflects multiple initiation sites or mRNA degradation. Suitable -10 and -35 regions for the guaBA operon were located by inspection and exhibit good homology with the consensus sequences. Comparison of the guaBA promoter and control region with either the purF (52) or purM (unpublished data) control regions does not reveal any common features. This is not unexpected since the guaBA operon is regulated differently from the *de novo* purine biosynthetic pathway (53). There are no obvious secondary structural features of the guaBA control region that would suggest that it is regulated by an attenuation mechanism as has been found for the pyrB (54) and pyrE (55) genes.

The regulatory mechanism of the guaBA operon is presently known to include repression by guanine nucleotides, induction by adenine nucleotides, a regulatory locus, guaR, and an uncharacterized class of unlinked regulatory mutations (16,17,18,20,46). The region of dyad symmetry surrounding the -10 region as well as that preceding the translation start codon could be involved in the regulation of the guaBA operon. The isolation, sequencing and characterization of guaBA control region regulatory mutations will allow us to determine if the dual adenine and guanine control mechanisms can act at separate regions.

One unexpected feature of the guaBA operon is the potential secondary structure present in the 68 bp intercistronic region (Fig. 5). This potential structure is not a member of the REP (56,57) sequences. This

structure has the features of a rho-independent terminator (51) and its possible role as an attenuator for the differential expression of the guaB and guaA gene products remains to be elucidated.

ACKNOWLEDGEMENTS

This work was supported by NIH Grant AI 20068. We thank B.A. Roe for invaluable advice on sequencing and Howard Zalkin for critical discussions and reading of the manuscript. The skillful typing of this manuscript by Carol Ann Ardoin is gratefully appreciated.

*To whom correspondence should be addressed

REFERENCES

1. Magasanik, B., Moyed, H.S. and Gehring, L.B. (1957) *J. Biol. Chem.* 236, 339-350.
2. Gilbert, H.S., Lowe, C.R. and Drabble, W.T. (1979) *Biochem. J.* 183, 481-494.
3. Krishnaiah, K.V. (1975) *Arch. Biochem. Biophys.* 170, 567-575.
4. Gilbert, H.S. and Drabble, W.T. (1980) *Biochem. J.* 191, 533-541.
5. Heyde, H. and Morrison, J.F. (1976) *Biochim. Biophys. Acta* 429, 635-644.
6. Wu, T.W. and Scrimgeour, K.G. (1973) *Can. J. Biochem.* 51, 1380-1390.
7. Wu, T.W. and Scrimgeour, K.G. (1973) *Can. J. Biochem.* 51, 1391-1397.
8. Hanon, W.H. and Levin, A.P. (1971) *Bacteriol. Proc.* 129,
9. Heyde, H., Nagabhushanam, M.V. and Morrison, J.F. (1976) *Biochim. Biophys. Acta* 429, 645-660.
10. Ishii, K. and Shiio, I. (1968) *J. Biochem.* 63, 661-669.
11. Buzzee, D.H. and Levin, A.P. (1968) *Biochem. Biophys. Res. Commun.* 30, 673-677.
12. Lambden, P.R. and Drabble, W.T. (1973) *J. Bacteriol.* 115, 992-1002.
13. Vales, L.D., Chase, J.W. and Murphy, J.B. (1979) *J. Bacteriol.* 139, 320-322.
14. Nijkamp, H.J.J. and Oskamp, A.A.G. (1968) *J. Mol. Biol.* 35, 103-109.
15. Fukumaki, Y., Shimada, K. and Takagi, Y. (1977) *J. Bacteriol.* 131, 685-688.
16. Nijkamp, H.J.J. and DeHaan, P.G. (1967) *Biochim. Biophys. Acta* 145, 31-40.
17. Mehra, R.K. and Drabble, W.T. (1981) *J. Gen. Micro.* 123, 27-37.
18. Nijkamp, H.J.J. (1969) *J. Bacteriol.* 100, 585-593.
19. Spibey, N. and Drabble, W.T. (1981) *J. Gen. Micro.* 126, 497-501.
20. Tiedeman, A.A. and Smith, J.M. (1984) *Mol. Gen. Genet.* 195, 77-82.
21. Miller, J.H. (1972) Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
22. Bachmann, B.J. (1972) *Bacteriol. Rev.* 36, 525-557.
23. Messing, J. (1979) *Recombinant DNA Technical Bulletin* 2, 43-48.
24. Neidhardt, F.C., Bloch, P.L. and Smith, D.F. (1974) *J. Bacteriol.* 119, 736-747.
25. Dagert, M. and Ehrlich, S.D. (1979) *Gene* 6, 23-28.
26. Clarke, L. and Carbon, J. (1976) *Cell* 9, 91-99.
27. Clarke, L. and Carbon, J. (1979) *Methods Enzymology* 68, 396-408.
28. Vieira, J. and Messing, J. (1982) *Gene* 19, 259-268.
29. Messing, J. (1983) *Methods Enzymology* 101, 20-78.

30. Schaffer, H.E. and Sederoff, R.R. (1981) *Anal. Biochem.* 115, 113-122.
31. Holmes, D.S. and Quigley, M. (1981) *Anal. Biochem.* 114, 193-197.
32. Thompson, J.A., Blakesley, K., Doran, K., Hough, C.J. and Wells, D. (1983) *Methods Enzymology* 100, 368-399.
33. Norgard, M.V. (1981) *Anal. Biochem.* 113, 34-42.
34. Guo, L., Yang, R.C.A. and Wu, R. (1983) *Nuc. Acids Res.* 11, 5521-5540.
35. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci.* 74, 5463-5467.
36. Messing, J. and Vieira, J. (1982) *Gene* 19, 269-276.
37. Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A.J.H., and Roe, B.A. (1980) *J. Mol. Biol.* 143, 161-178.
38. Larson, R. and Messing, J. (1983) *DNA* 2, 31-35.
39. Salser, W., Gesteland, R.F. and Bolle, A. (1967) *Nature* 215, 588-591.
40. Berk, A.J. and Sharp, P.A. (1978) *Proc. Natl. Acad. Sci.* 75, 1274-1278.
41. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) *Molecular Cloning*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
42. Sollner-Webb, B. and Reader, R.H. (1979) *Cell* 18, 485-499.
43. Eisenbeis, S.J. and Parker, J. (1981) *Mol. Gen. Genet.* 183, 115-122.
44. Shine, J. and Dalgarno, L. (1974) *Proc. Natl. Acad. Sci.* 71, 1342-1346.
45. Stormo, G.D., Schneider, T.D. and Gold, L.M. (1982) *Nuc. Acids Res.* 10, 2971-2996.
46. Thomas, M.S. and Drabble, W.T. (1984) *Mol. Gen. Genet.* 195, 238-245.
47. Casadaban, M.J., Chou, J. and Cohen, S.N. (1980) *J. Bacteriol.* 143, 971-980.
48. Grantham, R., Gautier, C., Gouy, M., Jacobzone, M. and Mercier, R. (1981) *Nuc. Acids Res.* 9, 43-47.
49. Grosjean, H. and Fiers, W. (1982) *Gene* 18, 199-209.
50. Shimada, K., Fukumaki, Y. and Takagi, Y., (1976) *Mol. Gen. Genet.* 147, 203-208.
51. Hawley, D.K. and McClure, W.R. (1983) *Nuc. Acids Res.* 11, 2237-2255.
52. Tso, J.Y., Zalkin, H., van Cleemput, M., Yanofsky, C. and Smith, J.M. (1982) *J. Biol. Chem.* 257, 3525-3531.
53. Gots, J.S., Benson, C.E., Jochimsen, B. and Koduri, K.R. (1977) in *Purine and Pyrimidine Metabolism*. Ciba Foundation Symposium 48, 23-41, Elsevier Press, Amsterdam.
54. Roof, W.F., Foltermann, K.F. and Wild, J. (1982) *Mol. Gen. Genet.* 187, 391-400.
55. Poulsen, P., Bonekamp, F. and Jensen, K.F. (1984) *EMBO J.* 3, 1783-1790.
56. Gilson, E., Clement, J.-M., Brutlag, D. and Hofnung, M. (1984) *EMBO J.* 3, 1417-1421.
57. Stern, M.J., Ames, G.F.-L., Smith, N.H., Robinson, E.C. and Higgins, C.F. (1984) *Cell* 37, 1015-1026.
58. Holmes, W.M., Platt, T. and Rosenberg, M. (1983) *Cell* 32, 1029-1032.
59. Landy, A. and Ross, W. (1977) *Science* 197, 1147-1160.
60. Pinkham, J.L., Platt, T., Enquist, L.W. and Weisberg, R.A. (1980) *J. Mol. Biol.* 144, 587-592.
61. Tinoco, I., Borer, P.N., Dengler, B., Levine, M.D., Uhlenbeck, O.C., Crothers, D.M. and Gralla, J. (1973) *Nature New Biology* 246, 40-41.