Volume 13 Number 19 1985 Nucleic Acids Research

DNA sequence and characterization of the *Escherichia coli serB* gene

Andrew F.Neuwald and George V.Stauffer

Department of Microbiology, University of Iowa, Iowa City, IA 52242, USA

Received 12 June 1985; Revised and Accepted 12 September 1985

ABSTRACT

We have determined the sequence of a DNA fragment containing the *Escherichia coli serB* gene. An open reading frame of 966 nucleotides was identified that encodes a polypeptide of 322 amino acids with a molecular weight of 35,002 daltons. The transcription start site was determined by Mung Bean nuclease mapping. The -10 and -35 regions of the *serB* promoter lack homology to the consensus sequences. In addition, the -35 region of the *serB* promoter overlaps the -35 region of a second divergent promoter. Frameshift mutations were constructed at three different sites within the *serB* gene. When plasmids carrying these mutations were used as templates in a minicell system, mutations closer to the proposed transcription and translation start sites resulted in smaller polypeptides than those further away, confirming the proposed direction of transcription and translation. The observed sizes of the truncated and native polypeptides were in agreement with those predicted from the DNA sequence. A very stable stem and loop structure (ΔG = -32 kcal/mole) that does not fit the criteria of known transcription terminators was found one nucleotide downstream from the putative UAA translation stop codon.

INTRODUCTION

The final step in serine biosynthesis in *Escherichia coli* is the hydrolysis of the phosphate group from phosphoserine to yield serine. This reaction is catalyzed by phosphoserine phosphatase (PSP) (1, 2), the *serB* gene product. The *serB* gene is located between the *deoD* and *trpR* loci, at about 100 minutes on the *E. coli* chromosome (3). A 5.3 kilobase pairs (kbp) fragment of the *E. coli* chromosome that contains the *serB* gene has been cloned and the approximate location of the *serB* gene within this fragment determined (4).

We were interested in sequencing and characterizing the *serB* gene for the following reasons. First, it is a constitutively expressed biosynthetic gene (1, 2) and may exhibit unique RNA polymerase and ribosome binding characteristics. Second, since the serine pathway is interrelated with a number of other metabolic pathways (5), characterization of the *serB* gene

may help reveal how *E. coli* orchestrates the numerous metabolic events in which serine plays a role. The present study serves as a step toward the understanding of the *serB* gene at the molecular level.

MATERIALS AND METHODS

Strains and plasmids

All bacterial strains are derivatives of *E. coli* K-12. Strain GS359 is F$^-$ *leu*B6, *thi*-2, *supE*, *lacY*1, *tonA*21, *hsdRS*, (*deoD*, *serB*, *trpR*, *dye*)Δ and was provided by R. Somerville. Strain GS200 is a minicell producer and was obtained from G. Zurawski. Plasmid pBR322 (6) was from G. Zurawski. Plasmid p*serB*59-1 contains the *serB* gene on a 5.3 kbp *Bam*HI fragment of *E. coli* chromosomal DNA cloned into pBR322 (4); it was obtained from R. Somerville. Other plasmids were isolated during this investigation.

Media

Luria broth (LB) and glucose minimal medium have been described previously (7). Supplements were added at the following concentrations: amino acids, 50 μg/ml (serine was added at 200 μg/ml); vitamins, 1 μg/ml; ampicillin (Ap), 150 μg/ml; chloramphenicol, 300 μg/ml for plasmid amplification.

Restriction, ligation and transformation

The procedures for restriction endonuclease cleavage, ligation and transformation have been described (7).

Plasmid isolation

Plasmid DNA was prepared from chloramphenicol-amplified cells (8) by a sodium dodecyl sulfate lysis procedure, followed by ethidium bromide-cesium chloride equilibrium density gradient centrifugation (9, 10). The quick screen method of Holmes and Quigley (11) was used for rapid analysis of small quantities of plasmid DNA.

DNA sequencing and nuclease mapping

The DNA sequencing procedure of Maxam and Gilbert was used (12) with the modifications of Smith and Calvo (13). Gel electrophoresis was carried out as described by Sanger and Coulson (14). DNA fragments used for sequencing were isolated from polyacrylamide gels as described previously (15). Fragments were labeled with ^{32}P at either the 5' end with T4 polynucleotide kinase or at the 3' end with DNA polymerase I (Klenow fragment) and were either strand separated or cleaved with a second restriction enzyme before analysis. Labeling and isolation procedures are described in Maniatis et al. (16).

The transcription initiation site for the *serB* gene was mapped as described previously (15) except that Mung Bean nuclease was used for single strand degradation instead of S1 nuclease. A 180 base pairs (bp) *Msp*I fragment containing the *serB* promoter was ^{32}P-labeled at the 5' ends with T4 polynucleotide kinase and the anticoding strand isolated electrophoretically. Cellular RNA was isolated from strain GS359 carrying plasmid pGS143, which contains the *serB* gene. About 1 µg of ^{32}P-labeled single-stranded DNA was hybridized to about 20 µg of total cellular RNA. The hybridization mixture was digested with 500 units of Mung Bean nuclease for 1 h at 30°C in 100 µl of Mung Bean buffer [0.03 M Na acetate (pH 4.6), 0.05 M NaCl, 1 mM ZnCl, and 5% glycerol]. Digestion products were electrophoresed adjacent to a DNA sequencing ladder of the same fragment.

The transcription initiation site for a second gene divergent from the *serB* gene was determined as described above except that a 275 bp *Bst*NI fragment that spans the promoter region was used for labeling.

In vitro mutagenesis

Plasmid pGS143 (5 µg), which contains the *serB* gene, was incubated with the restriction endonuclease *Msp*I (20 units) in the presence of ethidium bromide (EtBr) (0.030 mg/ml) for 40 minutes at 37°C. This procedure results in a single cleavage site per plasmid due to intercalation of EtBr upon relaxation of the supercoiled structure (17). The digestion mixture was loaded onto a 0.8% low melting temperature agarose gel and electrophoresed at 4°C. Linearized DNA was isolated from the gel as described (16). The recessed 3' ends created by cleavage of DNA with *Msp*I were filled in using dCTP, dGTP and DNA polymerase I (Klenow fragment) (16), and the DNA recircularized with T4 DNA ligase (Fig. 1). The recircularized plasmids, presumably having gained an *Sst*II site with concomitant loss of an *Msp*I site, were used to transform strain GS359 (Δ*serB*). Transformants were selected on LB + Ap plates and then spotted onto glucose minimal + Ap plates either with or without a serine supplement in order to detect insertion mutations within the *serB* gene. The approximate site of mutagenesis for each *serB* mutation was first determined by *Eco*RV + *Sst*II endonuclease digestion of plasmid DNA prepared by the quick screen method. Purified plasmid DNA was then prepared from three different *serB* mutants and digested with *Eco*RV + *Sst*II and *Msp*I to confirm that in each case an *Msp*I site had mutated to an *Sst*II site. Since the concomitant loss of an *Msp*I site and gain of an *Sst*II site is unlikely to occur except by the proposed mechanism (Fig. 1), the mutations were not confirmed by DNA sequencing.

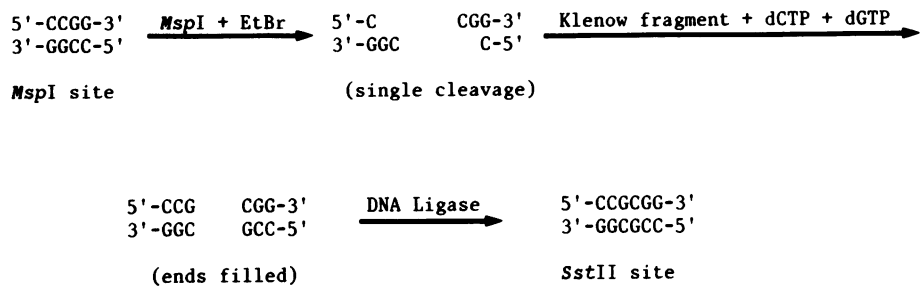

Figure 1. Scheme for the production of two base pair insertion mutations.

Minicell preparation

The minicell-producing strain GS200 was transformed with plasmid pGS143 or three derivatives of this plasmid (pGS154, pGS155 and pGS156) containing frameshift mutations in the *serB* gene. Growth and purification of minicells were as previously described (4). The procedure for ^{14}C-labeling and ^{35}S-labeling were identical except that either 0.2 μCi of ^{14}C-labeled amino acids or 10 μCi of [^{35}S]methionine was used (18). Labeled samples were electrophoresed on sodium dodecyl sulfate (SDS) polyacrylamide gels (15%) and the labeled polypeptides detected by autoradiography (18).

Chemicals and enzymes

Ethidium bromide, acrylamide, bisacrylamide and antibiotics were from Sigma Chemical Co. (St. Louis, MO); urea, ammonium persulfate and sequencing grade acrylamide were from Bio-Rad Laboratories (Richmond, CA); restriction endonucleases, T4 DNA ligase, DNA polymerase I (Klenow fragment) and bacterial alkaline phosphatase were purchased from Bethesda Research Laboratory (Gaithersburg, MD), or New England BioLabs (Beverly, MA); T4 polynucleotide kinase and Mung Bean nuclease were from P-L Biochemicals (Milwaukee, WI); [^{35}S]methionine, [^{14}C]amino acids, [γ-^{32}P]ATP and [α-^{32}P]dNTPs were from Amersham (Arlington Heights, IL); ^{14}C molecular weight standards were from New England Nuclear Corp. (Boston, MA); low melting temperature agarose was obtained from Marine Colloids BioProducts (Rockland, ME). All other chemicals were reagent grade and commercially available.

RESULTS

DNA sequencing and Mung Bean nuclease mapping

A restriction enzyme map of the 5.3 kbp *Bam*HI fragment containing the *serB* gene was constructed from analytical digestions prior to DNA sequencing

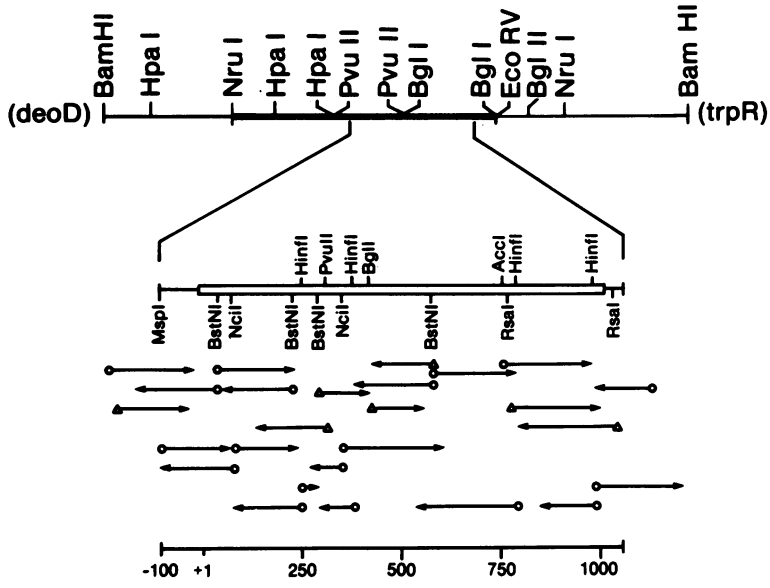

Figure 2. DNA sequencing strategy. A restriction enzyme map of the 5.3 kbp *Bam*HI fragment of the *E. coli* chromosome containing the *serB* gene is shown at the top. The orientation of this fragment with respect to the *deoD* and *trpR* region is indicated (R. Somerville, personal communication). The bold face indicates the 2.4 kbp *Nru*I-*Eco*RV fragment that was ligated into the *Nru*I-*Eco*RV sites of pBR322 to construct plasmid pGS143 (see Results). An expanded map of the region sequenced is shown below it. The open box indicates the *serB* gene. Restriction enzyme sites used for ^{32}P-labeling are given except for a 5'-labeled *Bst*NI site and a 3'-labeled *Pvu*II site (far left) and a 5'-labeled *Bst*NI site (far right). Open circles (o) indicate 5'-labeled ends and open triangles (Δ) indicate 3'-labeled ends. The locations of fragments relative to the transcription initiation site (+1) are given at the bottom of the figure.

(Fig. 2). It was previously reported that deletion of the two *Hpa*I fragments in the *Bam*HI insert of plasmid p*serB*59-1 abolished *serB* activity (4). A spontaneous deletion occurred during the initial removal of these fragments leading to the Ser$^-$ phenotype. Contrary to the original report, the positions of the *Hpa*I sites are as shown.

The DNA sequencing strategy is shown in Fig. 2. Both strands were sequenced for the entire region shown. The DNA sequence of the *serB* gene and the predicted amino acid sequence of PSP are shown in Fig. 3. Nucleotides are numbered from the transcription start site as determined from Mung Bean nuclease mapping (Fig. 4A). The -10 and -35 regions, as determined from the location of the prominent band, lack homology to the consensus sequences.

```
                                                                        TGAAAATTTGAGACT
                   GGC CTT AAA GTC AAA ACA CGC TCG GTA GAAGGTGGGAAACTTTTAAACTCTGA
                   Arg Phe Lys Leu Lys Thr Arg Ala Met     SD(+1)      (-10)

      -35                       -10      +1                             SD
TAATGTTGCCAGAAGCAATGGATACAAGGTAGCCTCATGCGTTATTTCCCTGCTTCGAACGATTTTACAGGAGCCTTA
ATTACAACGGTCTTCGTTACCTATGTTCCATCGGAGTACGCAATAAA
       (-35)

Met Pro Asn Ile Thr Trp Cys Asp Leu Pro Glu Asp Val Ser Leu Trp Pro Gly Leu Pro
ATG CCT AAC ATT ACC TGG TGC GAC CTG CCT GAA GAT GTC TCT TTA TGG CCG GGT CTG CCT

Leu Ser Leu Ser Gly Asp Glu Val Met Pro Leu Asp Tyr His Ala Gly Arg Ser Gly Trp
CTT TCA TTA AGT GGT GAT GAA GTG ATG CCA CTG GAT TAC CAC GCA GGT CGT AGC GGC TGG
        100
                                         50
Leu Leu Tyr Gly Arg Gly Leu Asp Lys Gln Arg Leu Thr Gln Tyr Gln Ser Lys Leu Gly
CTG CTG TAT GGT CGT GGG CTG GAT AAA CAA CGT CTG ACC CAA TAC CAG AGC AAA CTG GGT
                                                            200

Ala Ala Met Val Ile Val Ala Ala Trp Cys Val Glu Asp Tyr Gln Val Ile Arg Leu Ala
GCG GCG ATG GTG ATT GTT GCC GCC TGG TGC GTG GAA GAT TAT CAG GTG ATT CGT CTG GCA

                                                                            100
Gly Ser Leu Thr Ala Arg Ala Thr Arg Leu Ala His Glu Ala Gln Leu Asp Val Ala Pro
GGT TCA CTC ACC GCA CGG GCT ACA CGC CTG GCC CAC GAA GCG CAG CTG GAT GTC GCC CCG
                                    300

Leu Gly Lys Ile Pro His Leu Arg Thr Pro Gly Leu Leu Val Met Asp Met Asp Ser Thr
CTG GGG AAA ATC CCG CAC CTG CGC ACG CCG GGT TTG CTG GTG ATG GAT ATG GAC TCC ACC

Ala Ile Gln Ile Glu Cys Ile Asp Glu Ile Ala Lys Leu Ala Gly Thr Gly Glu Met Val
GCC ATC CAG ATT GAA TGT ATT GAT GAA ATT GCC AAA CTG GCC GGA ACG GGC GAG ATG GTG
        400
                                        150
Ala Glu Val Thr Glu Arg Ala Met Arg Gly Glu Leu Asp Phe Thr Ala Ser Leu Arg Ser
GCG GAA GTA ACC GAA CGG GCG ATG CGC GGC GAA CTC GAT TTT ACC GCC AGC CTG CGC AGC
                                                            500

Arg Val Ala Thr Leu Lys Gly Ala Asp Ala Asn Ile Leu Gln Gln Val Arg Glu Asn Leu
CGT GTG GCG ACG CTG AAA GGC GCT GAC GCC AAT ATT CTG CAA CAG GTG CGT GAA AAT CTG

                                                                            200
Pro Leu Met Pro Gly Leu Thr Gln Leu Val Leu Lys Leu Glu Thr Leu Gly Trp Lys Val
CCG CTG ATG CCA GGC TTA ACG CAA CTG GTG CTC AAG CTG GAA ACG CTG GGC TGG AAA GTG
                                    600

Ala Ile Ala Ser Gly Gly Phe Thr Phe Phe Ala Glu Tyr Leu Arg Asp Lys Leu Arg Leu
GCG ATT GCC TCC GGC GGC TTT ACT TTC TTT GCT GAA TAC CTG CGC GAC AAG CTG CGC CTG

Thr Ala Val Val Ala Asn Glu Leu Glu Ile Met Asp Gly Lys Phe Thr Gly Asn Val Ile
ACC GCC GTG GTA GCC AAT GAA CTG GAG ATC ATG GAC GGT AAA TTT ACC GGC AAT GTG ATC
        700
                                        250
Gly Asp Ile Val Asp Ala Gln Tyr Lys Ala Lys Thr Leu Thr Arg Leu Ala Gln Glu Tyr
GGC GAC ATC GTA GAC GCG CAG TAC AAA GCG AAA ACT CTG ACT CGC CTC GCG CAG GAG TAT
                                                            800

Glu Ile Pro Leu Ala Gln Thr Val Ala Ile Gly Asp Gly Ala Asn Asp Leu Pro Met Ile
GAA ATC CCG CTG GCG CAG ACC GTG GCG ATT GGC GAT GGA GCC AAT GAC CTG CCG ATG ATC

                                                                            300
Lys Ala Ala Gly Leu Gly Ile Ala Tyr His Ala Lys Pro Lys Val Asn Glu Lys Ala Glu
AAA GCG GCA GGG CTG GGG ATT GCC TAC CAT GCC AAG CCA AAA GTG AAT GAA AAG GCG GAA
                                    900

Val Thr Ile Arg His Ala Asp Leu Met Gly Val Phe Cys Ile Leu Ser Gly Ser Leu Asn
GTC ACC ATC CGT CAC GCT GAC CTG ATG GGG GTA TTC TGC ATC CTC TCA GGC AGC CTG AAT

Gln Lys end  —— ———►      ◄—————             SD            Met Ala Lys Ala
CAG AAG TAA TTGCTCGCCCGCCATCCTGCGGGCGGCACAGCATTAACGAGGTACACC GTG GCA AAA GCT
       1000
```

A second weakly protected band is visible in Fig. 4A eight nucleotides from the prominent band that may be a second site for transcription initiation. No Mung Bean nuclease resistant DNA was detected for the *serB* promoter when the mRNA for hybridization was isolated from a wild type *E. coli* not carrying plasmid pGS143 (data not shown). The predicted ATG initiation codon is preceded by a good Shine-Dalgarno sequence (19), followed by an open reading frame long enough to code for PSP (M_r = 35,000). A 10 bp hairpin structure, calculated to have a stability of ΔG = -32 kcal/mole (20), immediately follows the TAA stop codon at position +999 to +1001. Table 1 gives the codon usage data for PSP.

A transcription initiation site for a gene divergent from the *serB* gene was also detected by Mung Bean nuclease mapping (Fig. 4B). This initiation site is 62-63 bp upstream from the *serB* transcription initiation site, and its -10 and -35 regions show homology to the concensus sequences (Fig. 3). The -35 region of this promoter overlaps the -35 region of the *serB* promoter. A Shine-Dalgarno sequence and ATG start codon (Fig. 3) precede an open reading frame of 642 bp, and mutation analysis has verified that this region encodes a polypeptide with a molecular weight of about 25,000 daltons (unpublished).

Direction of transcription and translation of the *serB* gene

Plasmid pGS143 was constructed by ligation of the 2.4 kbp *Nru*I-*Eco*RV DNA fragment of plasmid p*serB*59-1 containing the *serB* gene (bold face in Fig. 2) into the *Nru*I-*Eco*RV sites of plasmid pBR322. This smaller plasmid facilitated the construction of 2 bp insertion mutations within the *serB* gene (see Methods).

Using plasmid pGS143, three insertion mutations were isolated that did not complement the *serB* deletion mutant GS359 (designated plasmids pGS154, pGS155, and pGS156). The location of each insertion was determined by restriction endonuclease mapping (Fig. 5). When these plasmids were used as templates in a minicell system, the polypeptides made from the *serB* gene were smaller than the *serB*$^+$ polypeptide (Fig. 6). Insertion mutations closer to the putative transcription initiation site resulted in smaller truncated

Figure 3. Nucleotide sequence of the *E. coli serB* gene and its flanking regions and the predicted amino acid sequence of PSP. The transcription initiation site determined by Mung Bean nuclease mapping is indicated by +1. The -10 and -35 regions and Shine-Dalgarno (SD) sequence are indicated. Arrows at the end of the *serB* gene indicate a region that can base pair to form a stem and loop structure. The Shine-Dalgarno sequences and deduced N-terminal amino acid sequences for a putative downstream gene and a putative upstream gene are also indicated. The +1, -10 and -35 regions of a divergent promoter that overlaps the *serB* promoter are indicated in parentheses. Only the coding strands are given for transcribed regions.

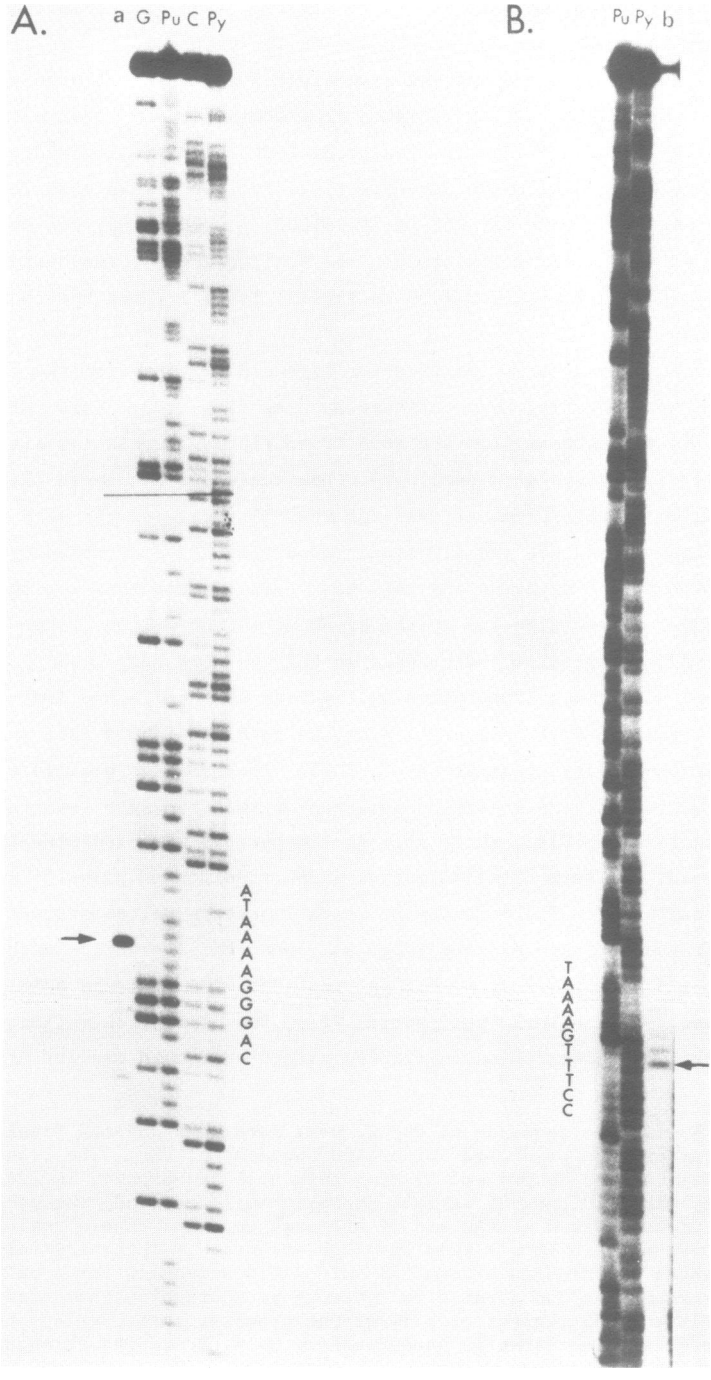
A.
a G Pu C Py
ATAAAAGGGAC
B.
Pu Py b
TAAAAGTTCC

Table 1. Codon Usage in *serB*

Phe	TTT	4	Ser	TCT	1	Tyr	TAT	3	Cys	TGT	1
Phe	TTC	2	Ser	TCC	2	Tyr	TAC	5	Cys	TGC	3
Leu	TTA	3	Ser	TCA	3	End	TAA	1	End	TGA	0
Leu	TTG	1	Ser	TCG	0	End	TAG	0	Trp	TGG	5
Leu	CTT	1	Pro	CCT	3	His	CAT	1	Arg	CGT	7
Leu	CTC	5	Pro	CCC	0	His	CAC	4	Arg	CGC	7
Leu	CTA	0	Pro	CCA	3	Gln	CAA	4	Arg	CGA	0
Leu	CTG	33	Pro	CCG	7	Gln	CAG	9	Arg	CGG	2
Ile	ATT	10	Thr	ACT	3	Asn	AAT	7	Ser	AGT	1
Ile	ATC	9	Thr	ACC	10	Asn	AAC	1	Ser	AGC	5
Ile	ATA	0	Thr	ACA	1	Lys	AAA	11	Arg	AGA	0
Met	ATG	11	Thr	ACG	5	Lys	AAG	5	Arg	AGG	0
Val	GTT	1	Ala	GCT	4	Asp	GAT	10	Gly	GGT	8
Val	GTC	3	Ala	GCC	15	Asp	GAC	9	Gly	GGC	12
Val	GTA	4	Ala	GCA	4	Glu	GAA	16	Gly	GGA	2
Val	GTG	14	Ala	GCG	14	Glu	GAG	3	Gly	GGG	5

polypeptides, confirming the direction of transcription and translation. The observed sizes of native and truncated *serB* polypeptides were in good agreement with the sizes predicted from the DNA sequence using the proposed ATG start codon (Table 2).

DISCUSSION

The nucleotide sequence of the *E. coli serB* gene has been determined. The open reading frame from nucleotide +33 to +998 (Fig. 3) was validated as coding for the *serB* gene based on the following criteria: (i) two base pair insertion mutations within the proposed *serB* gene resulted in a loss of complementation; (ii) a 35,000 dalton polypeptide, compatible with the 35,002 dalton protein coded by the open reading frame, disappeared in minicells when the *serB* mutant plasmids were used as templates (Fig. 6); (iii) frame-shift mutations closer to the proposed transcription and translation start sites resulted in smaller truncated polypeptides than those further away,

Figure 4. A. Mung Bean nuclease mapping of the 5' end of the *serB* gene. The 5' ^{32}P-labeled single stranded DNA fragment containing the *serB* regulatory region that was protected from nuclease digestion by hybridization to *serB* mRNA is shown in lane a. The site of transcription initiation was determined by comparison to a DNA sequence ladder of the same DNA fragment. B. Mung Bean nuclease mapping of the 5' end of a second gene divergent from the *serB* gene. The single stranded DNA fragment protected from nuclease digestion by hybridization to mRNA is shown in lane b.

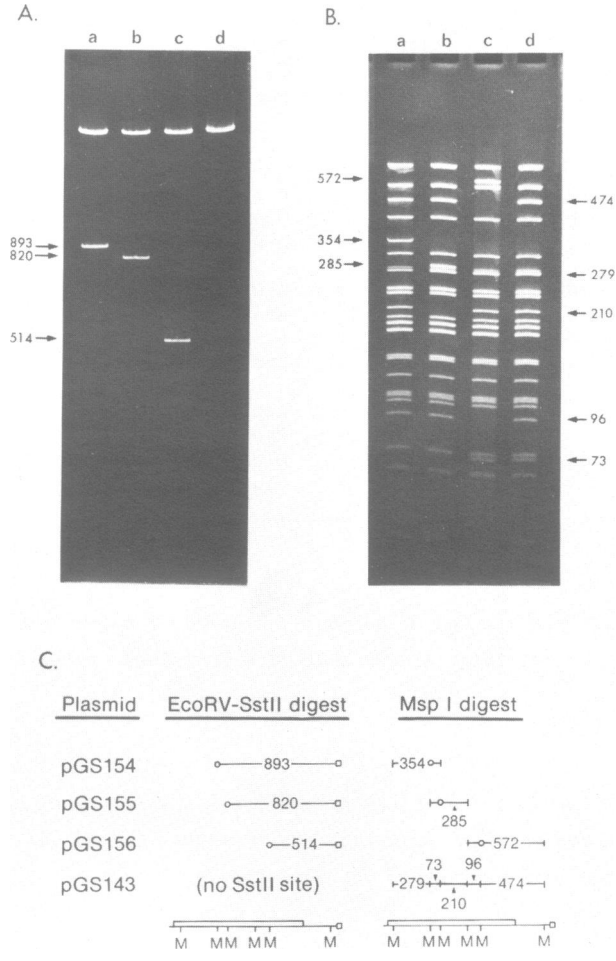

Figure 5. Location of 2 bp insertion mutations in the *serB* gene by restriction enzyme mapping. A. *Eco*RV-*Sst*II digests of plasmid pGS143 and derivatives that contain 2 bp insertions which generate unique *Sst*II sites within the *serB* gene (3.5% polyacrylamide gel). The arrows indicate the locations of the diagnostic *Eco*RV-*Sst*II fragments. The sizes of the fragments indicate the locations of the *Sst*II sites within the *serB* gene (see Fig. 5C). Lanes: a, pGS154; b, pGS155, c. pGS156; d, pGS143. B. *Msp*I digests of plasmid pGS143 and derivatives that contain 2 bp insertions in the *serB* gene (8% polyacrylamide gel). Arrows on the right indicate the locations of the *Msp*I fragments that became fused in the mutant plasmids. Arrows on the left indicate the locations of the new *Msp*I fragments generated by the 2 bp insertions (see Fig. 5C). Lanes: a, pGS154; b, pGS155, c, pGS156; d, pGS143. There are two 279 bp *Msp*I fragments in pGS143, one of which became fused in pGS154. C. Locations of *Eco*RV-*Sst*II and *Msp*I fragments as determined by restriction endonuclease digestion and DNA sequence data.

Plasmid designations are given in the left column. Diagnostic *Eco*RV-*Sst*II and *Msp*I fragments (see Fig. 5A and 5B) are shown to the center and right respectively. All of the *Msp*I fragments within the region of interest are shown for the parental plasmid whereas only the fused fragments are shown for mutant plasmids. Sizes of fragments are given in base pairs. Symbols: *Eco*RV site, open square (□); *Sst*II site, open circle (o); *Msp*I site, M. The open box indicates the location of the *serB* structural gene.

confirming the proposed direction of transcription and translation; (iv) the sizes of the truncated polypeptides produced in these frameshift mutants corresponded to the predicted sizes as determined from the nucleotide sequence (Table 2).

A 32-bp leader sequence precedes the proposed ATG translation initiation codon. This codon is preceded by a sequence complementary to the 3' end of 16S rRNA (Shine-Dalgarno sequence), and is at the beginning of an open reading frame that codes for a protein of *Mr* 35,002. The next closest possible start codons are a GTG codon at position +114 to +116 and an ATG codon at position +117 to +119. Both would produce polypeptides with molecular weights about 3,000 daltons smaller than those initiated from the proposed site, and both lack a good Shine-Dalgarno sequence. These results suggest that the proposed translation initiation site is correct.

The -10 and -35 regions of the *serB* promoter lack homology to the consensus sequences. Although sequences are present that show homology to the consensus -10 (TATAAT) and -35 (TTGACA) sequences, they are inappropriately spaced from the transcription start site determined by Mung Bean nuclease mapping. This is difficult to explain at this time. Perhaps there is another sigma factor produced by *E. coli* that can recognize the *serB* promoter. An additional sigma factor (σ^{32}) is involved in the heat-shock response of *E. coli* (23). Interestingly, the *serB* -10 region (CCTCATGC) shows considerable homology to the σ^{32} consensus sequence (CCCCATTT) (24). However, an *E. coli* strain carrying a plasmid with a *serB-lacZ* fusion did not produce increased levels of β-galactosidase after heat shock at 42°C for 5 min (data not shown). Perhaps σ^{32} is involved in the transcription of the *serB* gene independent of the heat shock response. Alternatively, the *serB* gene may have a poorly transcribed but very stable messenger RNA. Some poorly transcribed genes, such as the *malT* and *lacI* genes, have either promoter -10 or -35 regions with poor homology to the consensus sequences (25, 26, 27). Additional experiments, such as measurement of *serB* mRNA synthesis, mRNA stability studies and replacement of nucleotides within the promoter region will help to resolve these questions.

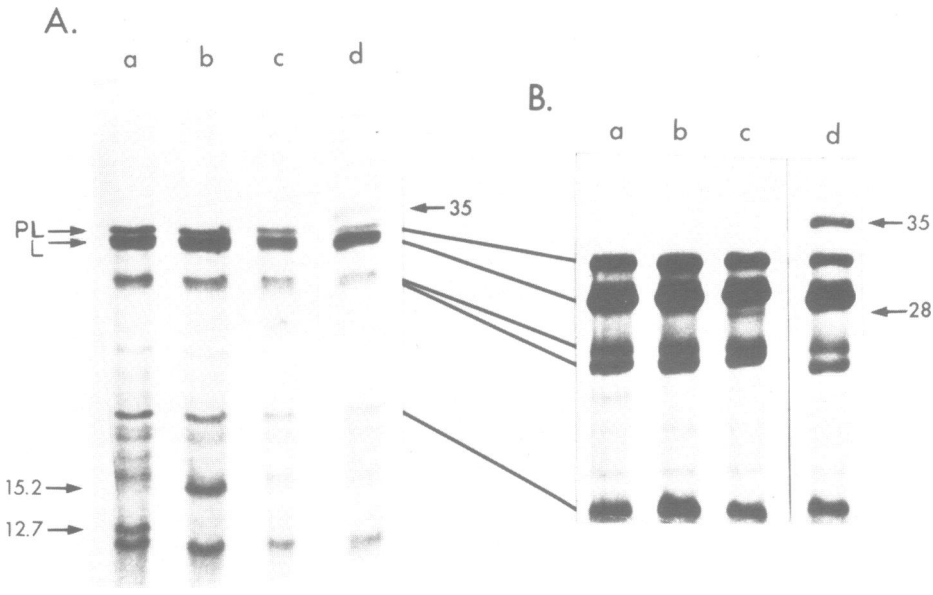

Figure 6. SDS polyacrylamide gels (15%) of truncated and native *serB* polypeptides synthesized and labeled in a minicell system. A. Autoradiograph of ^{14}C-labeled polypeptides. B. Autoradiograph of [^{35}S]-methionine labeled polypeptides. The locations of truncated and native polypeptides are indicated by arrows. The sizes are in kilodaltons. The truncated *serB* polypeptide produced with plasmid pGS156 as template could only be detected using ^{35}S-methionine. Templates in both A and B are: lane a, pGS154; lane b, pGS155; lane c, pGS156; lane d, pGS143. The molecular weight markers (not shown) include cytochrome c (*Mr*, 12,300), lactoglobin A (*Mr*, 18,357), ovalbumin (*Mr* 46,000). PL and L indicate the precursor (*Mr* 31,400) and processed (*Mr* 28,800) forms respectively of plasmid encoded β-lactamase. The molecular weights of β-lactamase standards were determined from Sutcliffe (21). Three plasmid encoded polypeptides (*Mr* 25,000, *Mr* 19,000 and *Mr* 12,000) as well as a β-lactamase degradation product (*Mr* 25,000) (22) are also evident but for clarity are not labeled on the figure.

The *serB* promoter overlaps in its -35 region with a second divergent promoter (Fig. 3). The genes coding for the functionally related *Pst*I restriction and modification enzymes have previously been shown to have overlapping divergent promoters (28). Perhaps there is a functional relationship between the *serB* gene and the divergent gene.

The very stable GC rich stem and loop structure at the 3' end of the *serB* gene may be a transcription terminator. However, the lack of any T residues immediately downstream from the stem and loop structure excludes this as a rho independent terminator and its close proximity to the *serB*

Table 2. Molecular weights of native and truncated *serB* polypeptides

Plasmid	Site of mutagenesis	Predicted C-terminal amino acid sequence[a/]	Size of polypeptide[b/] Calculated	Observed
pGS154	Pro 110 codon	Pro Arg Val Cys Trp end CC<u>G</u> <u>C</u>GG GTT TGC TGG TGA	12,682	12,700
pGS155	Ala 134 codon	Ala Ala Glu Arg Ala Arg Trp Trp Arg Arg end GCC <u>GC</u>G GAA CGG GCG AGA TGG TGG CGG AGG TAA	15,949	15,200
pGS156	Thr 236 codon	Thr Ala Ala Met end ACC <u>GC</u>G GCA ATG TGA	26,154	28,000
pGS143	none		35,002	35,000

a/ The 2 bp insertions that generate the out of phase translation termination codons are underlined.

b/ The calculated and observed sizes of polypeptides (in daltons) are determined from the DNA sequence data (Fig. 3) and SDS-polyacrylamide gels (Fig. 6), respectively.

translation stop codon (one nucleotide), its GC rich stem, and a significant amount of potential upstream RNA secondary structure (not shown) suggest that it does not serve as a rho-dependent terminator (29). Perhaps it has some other function not related to transcription termination, such as imparting a greater stability to the mRNA transcript.

In addition, there is a good Shine-Dalgarno sequence 6 bp downstream from this hairpin structure, followed by a possible GTG start codon (Fig. 3). The start codon is followed by an open reading frame of at least 480 bp (unpublished). Therefore, it is possible that the *serB* gene is part of an operon that codes for at least one other gene. This gene may correspond to a 55,000 dalton protein which was previously associated with the region downstream from *serB* (4). We are presently characterizing this region to determine if there is any relationship between this gene and the *serB* gene.

The production of two base pair insertion mutations by *Msp*I-*Sst*II mutagenesis is a better method for the determination of the direction of transcription and translation and the approximate translation start site than the previously described Tn*5* insertion technique (30). Tn*5* insertional mutagenesis results in a larger plasmid and in a greater number of plasmid encoded polypeptides. Both of these factors hinder the characterization of these mutants. *Msp*I-*Sst*II insertional mutagenesis avoids these problems as well as allowing a more accurate localization of the site of mutagenesis. The concomitant loss of an *Msp*I site and gain of an *Sst*II site eliminates the need to verify the site of mutagenesis by DNA sequencing. The

site of translation termination for truncated polypeptides can be precisely determined if the nucleotide sequence is known. One possible drawback to this technique is that some *Msp*I sites may be preferentially cut over other sites, making it difficult to isolate mutations at a number of locations within a particular gene. However, this problem may be circumvented by using the same technique with other restriction enzymes, such as *Taq*I and *Hin*PI to generate *Nru*I and *Bss*HII sites, respectively. Use of these restriction enzymes will result in 2-bp frameshift mutations.

ACKNOWLEDGEMENTS

This investigation was supported by Public Health Service grant GM 26878 from the National Institute of General Medical Sciences.

REFERENCES

1. Pizer, L. I. (1963) J. Biol. Chem. 238, 3934-3944.
2. Umbarger, H. E., Umbarger, M. A. and Siu, P. M. L. (1963) J. Bacteriol. 85, 1431-1439.
3. Bachmann, B. J. (1983) Microbiol. Rev. 47, 180-230.
4. Garnant, M. K. and Stauffer, G. V. (1984) Mol. Gen. Genet. 193, 72-75.
5. Stauffer, G. V. (1983) in Herrmann, K. M. and Somerville, R. L. (eds.), Amino Acids: Biosynthesis and Genetic Regulation, Addison-Wesley, Reading, MA, pp. 103-113.
6. Bolivar, F., Rodriguez, R. L., Greene, P. J., Betlach, M. C., Heyneker, H. L. and Boyer, H. W. (1977) Gene 2, 95-113.
7. Stauffer, G. V., Plamann, M. D. and Stauffer, L. T. (1981) Gene 14, 63-72.
8. Clewell, D. B. (1972) J. Bacteriol. 110, 667-676.
9. Guerry, P., LeBlanc, D. J. and Falkow, S. (1973) J. Bacteriol. 116, 1064-1066.
10. Selker, E., Brown, K. and Yanofsky, C. (1977) J. Bacteriol. 129, 388-394.
11. Holmes, D. S. and Quigley, M. (1981) Analyt. Biochem. 114, 193-197.
12. Maxam, A. M. and Gilbert, W. (1980) Meth. Enzymol. 65, 449-560.
13. Smith, D. R. and Calvo, J. M. (1980) Nucl. Acids Res. 8, 2255-2274.
14. Sanger, F. and Coulson, A. R. (1978) FEBS Lett. 87, 107-110.
15. Plamann, M. D. and Stauffer, G. V. (1983) Gene 22, 9-18.
16. Maniatis, T., Fritsch, E. F. and Sambrook, J. (1982) Molecular Cloning. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
17. Parker, R. C., Watson, R. M. and Winograd, J. (1977) Proc. Natl. Acad. Sci. USA 74, 851-855.
18. Miozzari, G. F. and Yanofsky, C. (1978) J. Bacteriol. 133, 1457-1466.
19. Shine, J. and Dalgarno, L. (1974) Proc. Natl. Acad. Sci. USA 71, 1342-1346.
20. Tinoco, I., Jr., Borer, P. N., Dengler, B., Levine, M. D., Uhlenbeck, O. C., Crothers, D. M. and Gralla, J. (1973) Nature New Biol. (London) 246, 40-41.
21. Sutcliffe, J. G. (1978) Proc. Natl. Acad. Sci. USA 75, 3737-3741.
22. Dougan, G., Saul, M., Twigg, A., Gill, R. and Sherratt, D. (1979) J. Bacteriol. 138, 48-54.

23. Grossman, A. D., Erickson, J. W. and Gross, C. A. (1984) Cell 38, 383-390.
24. Cowing, D. W., Bardwell, J. C. A., Craig, E. A., Woolford, C., Hendrix, R. W. and Gross, C. A. (1985) Proc. Natl. Acad. Sci. USA 82, 2679-2683.
25. Hawley, D. K. and McClure, W. R. (1983) Nucl. Acids Res. 11, 2237-2255.
26. Chapon, C. (1982) EMBO J. 1, 369-374.
27. Calos, M. (1978) Nature 274, 762-765.
28. Walder, R. Y., Walder, J. A. and Donelson, J. E. (1984) J. Biol. Chem. 259, 8015-8026.
29. von Hippel, P. H., Bear, D. G., Morgan, W. D. and McSwiggen, J. A. (1984) Ann. Rev. Biochem. 53, 389-446.
30. Schulte, L. L., Stauffer, L. T. and Stauffer, G. V. (1984) J. Bacteriol. 158, 928-933.