Volume 17 Number 8 1989 Nucleic Acids Research

Effect of nucleotide sequences directly downstream from the AUG on the expression of bovine somatotropin in *E.coli*

C.-S.C.Tomich*, E.R.Olson, M.K.Olsen, P.S.Kaytes, S.K.Rockenbach and N.T.Hatzenbuhler[1]

Molecular Biology Research and [1]Biopolymer Chemistry, The Upjohn Company, Kalamazoo, MI 49007, USA

Received December 1, 1988; Revised and Accepted March 13, 1989

ABSTRACT

We have studied the expression of bovine somatotropin (BSt) to gain more understanding of various factors affecting translation in *E. coli*. The unmodified cDNA coding for mature bovine somatotropin does not produce significant amounts of BSt in *E. coli* using a pBR322-derived vector. However, a translation fusion with 16 codons from trpLE in front of BSt cDNA results in greater than 20% of total cell protein as the fusion product. Analysis of transcription by measuring the rate and integrity of the mRNA confirms that a post-transcriptional event is responsible for the poor expression of the BSt cDNA. There are two potential stem-loop structures in the 5' region of the mRNA which may interfere with translation. To study their effect on translation, lacZ fusions and oligonucleotide mutagenesis were carried out. The results demonstrate that the secondary structure involving the initiation codon blocks translation initiation. Removal of this stem-loop results in a 100-fold increase in BSt expression. However, the expression level is still low, amounting to only 0.5-1% of total cell protein. High level expression can be obtained by replacement of the beginning sequence of BSt cDNA with trpLE codons. These results suggest that in addition to the secondary structure, the nucleotide sequence or amino acid context within the beginning of BSt is incompatible with one of the steps in translation initiation.

INTRODUCTION

The possible use of animal somatotropins in agriculture has promoted several industrial laboratories to develop bacterial strains and vectors for efficient production of the proteins. Several strategies have been employed to obtain high level expression of BSt in *E. coli*, including changing the sequence around the Shine-Dalgarno site, altering the sequence in the 5' coding region of the gene, expressing the gene as the second part of a two cistron arrangement, or using high copy number plasmids (1-4; John Mott, personal communication). All of these approaches apparently overcome a step in translation initiation which is rate limiting in the unmodified cDNA. These studies suggest that one of the factors involved in limiting translation initiation may be the presence of secondary structure in the mRNA which would block access by the initiation complex. However, it is not

clear what structure, if any, is involved or if its elimination is indeed responsible for improving expression. By understanding in more detail what changes in the beginning of the cDNA affect expression, we hope to gain more insight into translation initiation in *E. coli*. In this study, we describe experiments demonstrating that a weak stem-loop structure obscuring the initiation codon AUG is a critical factor responsible for poor expression of the BSt cDNA. In addition to the secondary structure, the first few codons in the BSt gene either interferes with or is lacking the component to interact with the translation initiation steps for optimal expression.

MATERIALS AND METHODS

Enzymes, Plasmid Constructions and Oligonucleotides

Restriction endonucleases, T4 DNA ligase, and T4 polynucleotide kinase were purchased from New England Biolabs and Klenow fragment of *E. coli* DNA polymerase I was from Boehringer Mannheim. Enzyme reaction conditions were according to that of the supplier.

Plasmids were constructed and screened in *E. coli* HB101 or DH1. Transformation, isolation of DNA fragments, small and large scale plasmid preparations, and other basic techniques were according to that described by Maniatis et al. (5). Plasmids containing the trp promoter were grown in the presence of 100-200 μg/ml of tryptophan to repress the promoter. Plasmid constructs were analyzed with restriction endonucleases and by sequencing using the dideoxy chain termination method on double-stranded templates (6).

Oligonucleotides used were synthesized on an ABI 380B DNA Synthesizer, purified by electrophoresis on polyacrylamide gel with urea, and desalted with a SEP-PAK C18 cartridge from Waters Associates. To avoid insertion of more than one set of oligonucleotides, the oligonucleotides were not phosphorylated. When only 2 complementary oligonucleotides were used for insertion into a vector, the oligonucleotides were annealed first as described below and then used directly in ligation. When 4 or more oligonucleotides were needed, they were ligated and isolated as follows. Only the 5'-OH ends of those oligonucleotides to be ligated together were phosphorylated with T4 polynucleotide kinase in the presence of [γ-^{32}P]ATP followed by a chase with an excess of unlabelled ATP. After inactivating the kinase by boiling, the other oligonucleotides were added for annealing. Annealing was performed in 50 mM Tris-HCl, pH 7.8 and 10 mM $MgCl_2$ by boiling for about 5 min, slow cooling to room temperature over a period of 2-4 hours, and sometimes slow cooling to 4°C. The annealed mixture was ligated

and separated by polyacrylamide gel electrophoresis. The correctly assembled oligonucleotide block, visualized by autoradiography, was recovered by electro-elution, extracted with phenol, and precipitated with ethanol.

Induction and Analysis of BSt Expression

E. coli K12 (ATCC #e23716) was transformed with the various BSt plasmids and analyzed for expression. All cultures were grown at 37°C with aeration. All overnight cultures were in Luria broth with 0.2% glucose, 100 μg/ml ampicillin and 100-200 μg/ml tryptophan (LB). Induction was done in M9 minimal salts medium supplemented with 0.2% glucose, 100 μg/ml ampicillin and 0.05-0.2% autoclaved acid-hydrolyzed casein amino acids (M9CA). After a 50 to 100-fold dilution of the overnight culture into M9CA, the cells were grown to A_{550} of 0.3 to 0.4, harvested by centrifugation, and stored as a frozen cell pellet or sonicated cell extract. In cultures where cell growth was slow due to high level expression of BSt, the induction was carried out by collecting cells grown in LB to A_{550} of 0.5-0.7, washing in M9CA, resuspending in M9CA to A_{550} of 0.5, and allowing to grow for 2-4 hours.

The amount of BSt in a culture was analyzed by SDS polyacrylamide gel electrophoresis (SDS-PAGE), by Western immunoblotting (7), or by radio-immuno assay (R. Olsen, personal communication). Western blots were carried out by transferring the protein to nitrocellulose filter followed by reaction with rabbit anti-BSt polyclonal antibody and protein A-HRP as described by the supplier (BioRad). Sonicated extracts were prepared in 10 mM Tris-HCl and 1 mM EDTA. Total protein concentrations were determined using an assay kit (based on binding to Coomassie Blue) from BioRad. SDS-PAGE of whole cell extracts was by a gel system where the cross linker is N,N'-diallyltartar diamide (8), at an acrylamide concentration of 17% in the separating gel. Molecular weight standards were purchased from BioRad. Densitometric scanning of protein gels stained with Coomassie Blue was done with a Shimadzu Scanner.

Induction of cultures for β-galactosidase and preparation of extracts were done as described above for BSt expression. The host strain E405 was constructed by P1 transduction of K37 (9) to *lacΔ*, *proC::Tn10* using P1 grown on WPS18 (obtained from Don Court, Bionetics Research Institute; the *lac* deletion is described in reference 10). Sonicated extracts were prepared as described above and used for assay. Levels of β-galactosidase are expressed as Miller units (11).

Measurement of mRNA

The level of BSt-specific mRNA was measured by RNA pulse-labeling and RNA/DNA hybridization, Northern analysis, and primer extension analysis. Pulse-labeling conditions were modifications of that of Court et al. (12) and Bogosian et al. (13). Ten or 20 ml cultures were labeled with 25 μCi/ml of [5-^{3}H]uridine at 37°C for various periods of time, killed by the addition of cold buffer with NaN_3 and chloramphenicol, lysed with lysozyme and freeze-thawed, treated with DNase, extracted with hot phenol, and total RNA was precipitated with ethanol. To determine the amount of pulse-labeled mRNA, RNA/DNA hybridization was carried out according to the procedures of Baughman and Nomura (14) by baking the single-stranded DNA probe onto nitrocellulose filters for hybridization with the RNA in buffer containing formamide, SDS and carrier tRNA. The DNA probe was a PvuII fragment containing 85% of the BSt coding sequence (codons 24-188) cloned into the SmaI and the HincII site of the M13 vector mp8 according to the procedures of Messing (15). The mp8 containing the coding strand for BSt was used to measure BSt-specific mRNA and the one containing the anti-sense strand was used to measure background hybridization which was detected at very low levels.

Northern analysis was according to the procedures described by Maniatis et al. (5). The RNA was prepared as described above from cultures incubated in the presence of rifampicin for various times, electrophoresed in 1.4% agarose gel containing formaldehyde, and probed with a PvuII fragment containing 85% of the BSt sequence (codons 24-188). The probe was labeled with [γ-^{32}P]dCTP by nick translation. To serve as a control, a HinfI-EcoRI fragment containing β-lactamase sequence (16) was used as a probe.

Primer extension was carried out essentially according to the procedures of Polites and Marotti (17) for the first strand synthesis of cDNA, using an oligonucleotide complimentary to codons 63-69 of the BSt coding sequence. Reverse transcriptase and dNTPs were used to extend the oligonucleotide to the 5′ end of the mRNA. A dideoxy sequencing ladder of the BSt plasmid using the same oligonucleotide as a primer was run to serve as a marker for the 5′ end of the mRNA.

Determination of Plasmid Copy Number

E. coli cells containing plasmid were grown under tryptophan starvation conditions for trp promoter induction (see section on Induction and Analysis of BSt Expression) and labeled with 10 μCi/ml of [methyl-^{3}H]thymidine (25 Ci/mmol). Cells were collected by centrifugation, lysed with lysozyme

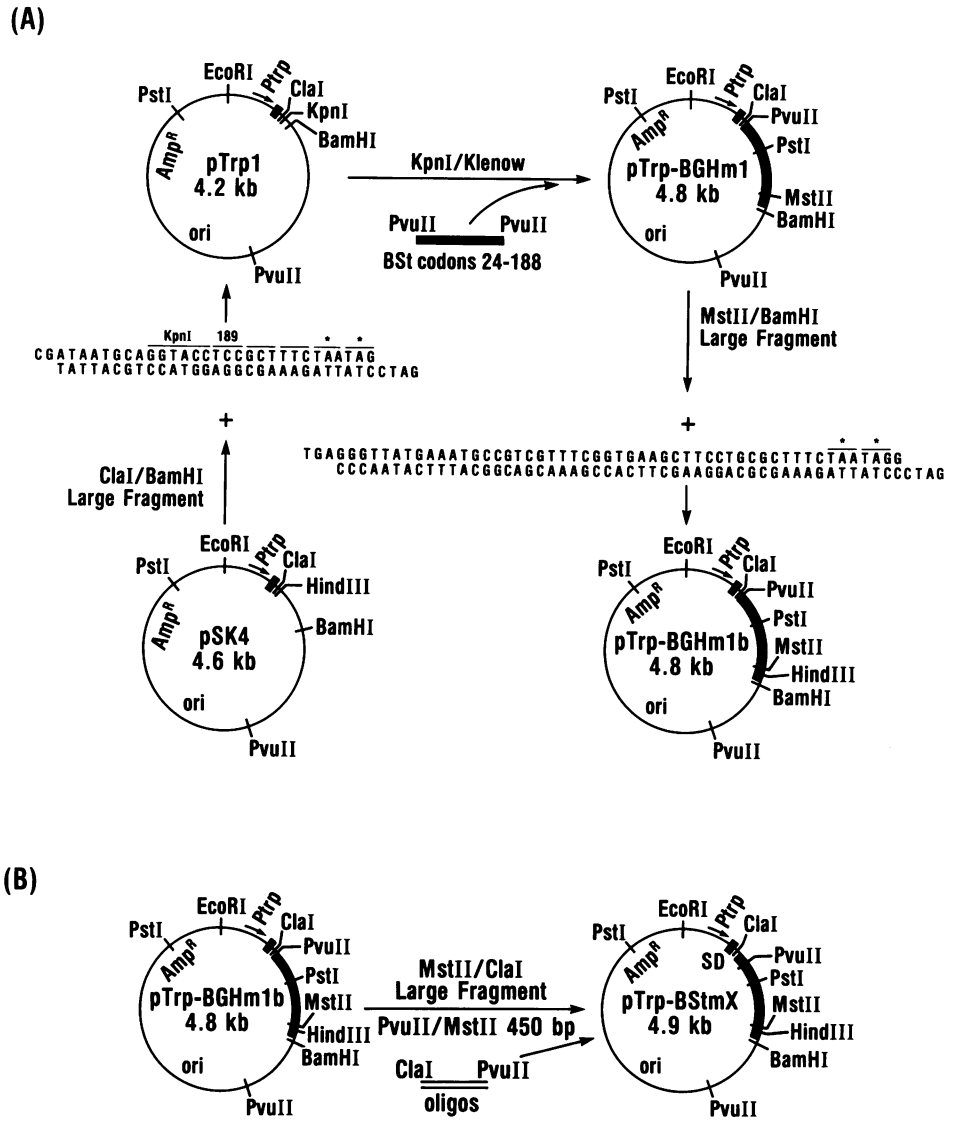

Fig. 1 (A) Construction of the vector pTrp-mlb.(B) Strategy used to make alterations in the beginning of the BSt gene.

followed by Brij 58 and deoxycholate, treated with RNase A and subjected to CsCl-ethidium bromide equilibrium gradient centrifugation. The gradient was fractionated onto filter papers which were then dried, washed with 5% TCA and measured for radioactivity. Copy number was calculated from the

radioactivity in the fractions containing the plasmid and the fractions containing chromosomal DNA, normalized for the size of the plasmid and the chromosome which is taken as 4,000 kb.

Calculation of Secondary Structure

To find the secondary structures in mRNA and calculate their free energy values, a computer program (MicroGenie from Beckman) based on the rules of Queen and Korn (18) was used. These calculations give the free energy values, in kcal, for the first stem-loop in pSK102, BSt-m4, BSt-m5, BSt-m6, BSt-m8, BSt-m9 and BSt-m10 of -6.2, -2.8, -2.0, -9.2, -6.8, -2.8 and -10.2, respectively. These values are used in this study. Using Tinoco rules (19), the values for pSK102, BSt-m4, BSt-m5, BSt-m6, BSt-m8, BSt-m9 and BSt-m10 are -10.0, -2.8, -2.8, -10.0, -6.8, -2.8 and -10.0, respectively, while the Salser rules (20) give -8.5, -3.2, -3.2, -8.5, -8.4, -4.9 and -10.2, respectively. Although the absolute number for a given stem-loop varies with the different rules, it does not affect the qualitative comparison of these stem-loops in our study.

Construction of A Vector for Easy Manipulation of the Beginning of BSt Gene

To test expression from BSt sequences with modifications at the beginning of the gene, the vector pTrp-BGHmlb was made to allow replacements of the first 23 codons for BSt with oligonucleotides. As shown in Fig. 1A, expression vector pSK4 (21) was modified to give pTrpl by replacement of the small ClaI-BamHI region with 2 oligonucleotides which contain a KpnI site upstream of the BSt codons 189-191 (codon 189 changed from cysteine to serine) and two stop codons. A 494 bp PvuII fragment containing codons 24 to 188 of the BSt cDNA was inserted into the KpnI site (single-stranded ends removed by Klenow enzyme) of pTrpl to give pTrp-BGHml. To remove the PvuII site at codon 188 of BSt, the small MstII to BamHI region (codon 176 to the last codon 191 of BSt) in pTrp-BGHml was replaced by 2 oligonucleotides (containing a HindIII site in the middle of the BSt codons 176 to 191 and two stop codons) to yield pTrp-BGHmlb. In pTrp-BGHmlb the BSt sequence starts at the PvuII site and therefore is truncated for the first 23 codons. It contains BSt codons 24-191 followed by two stop codons with the region for codons 176-191 (downstream of MstII) modified to optimize codon usage and to insert a HindIII site. This vector pTrp-BGHmlb was used for constructing all the plasmids with sequence modifications in the beginning of the BSt gene. To make those plasmids (Fig. 1B), oligonucleotides containing various modifications in the first 23 codons of the BSt gene can be inserted between ClaI and PvuII upstream of the truncated BSt sequence in

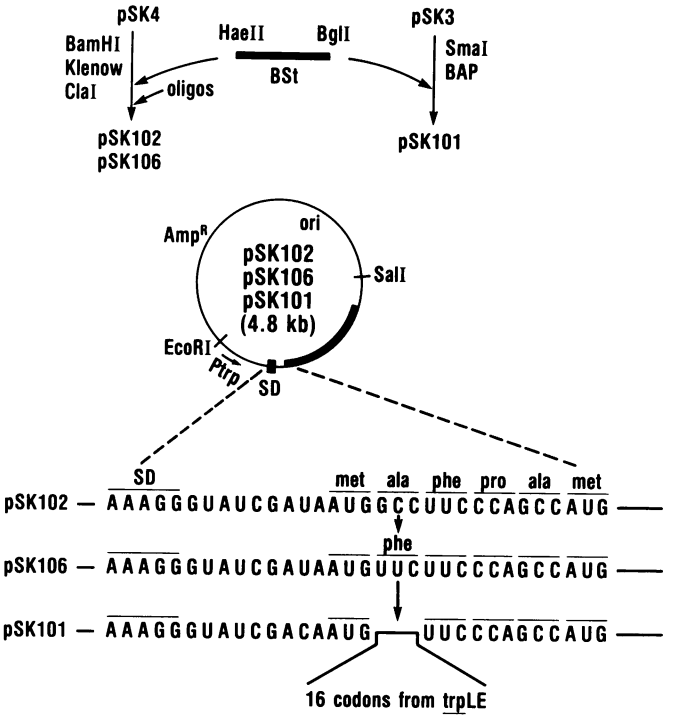

Fig. 2 Construction of pSK102, pSK106 and pSK101. The differences in these plasmids (the sequence immediately after the initiation codon AUG) are shown.

pTrp-BGHmlb. Due to another PvuII site in the vector, the insertion of these oligonucleotides involves separating pTrp-BGHmlb into 2 parts, a 450 bp PvuII-MstII fragment and the large MstII-ClaI fragment. The resulting plasmids, represented as pTrp-BStmX, contain modifications in the beginning of the BSt gene under control of the trp promoter and the trpL ribosome binding site.

RESULTS

Lack of BSt Expression from Unmodified cDNA

To establish a vector for the expression of BSt, the unmodified cDNA for BSt (obtained from Fritz Rottman, Case Western Reserve University) was placed in a pBR322-derived expression vector. As shown in Fig. 2, the ca. 800 bp HaeII-BglI fragment containing the cDNA sequence coding for the entire mature BSt, except for the first codon (alanine) was inserted into the expression vector pSK4 (21) behind the trp promoter and a ribosome

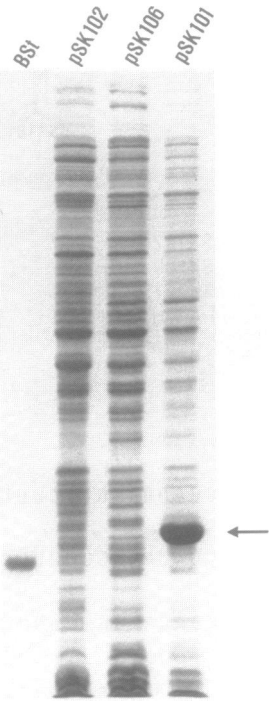

Fig. 3 Coomassie stained SDS-polyacrylamide gel of *E. coli* extracts containing the BSt expression plasmids pSK102, pSK106 and pSK101.The purified BSt is used as a molecular standard. The arrow indicates the band from the fusion of BSt to 16 trpLE codons in pSK101.

binding site essentially identical to that of trpL in a pBR322 background. Oligonucleotides were used to supply the missing first codon, the initiation codon ATG and the ClaI site sequence for cloning into the vector. The resulting plasmid, pSK102, directs less than 0.01% of total cell protein as BSt. This low level of BSt was barely detectable in a radio-immuno assay but was undetectable in Western analysis. Since the expression vector pSK4 can direct the expression of human renin cDNA to high levels (21), the lack of BSt expression is likely due to some property in the cDNA sequence itself or in its context in this vector.

To test if the expression could be increased by alterations in the beginning of the BSt sequence, plasmids pSK106 and pSK101 were constructed. As shown in Fig. 2, plasmid pSK106 was made in a manner identical to that for pSK102, i.e., placing the HaeII-BglI fragment containing BSt cDNA and oligo-nucleotides into pSK4. Plasmid pSK101 was constructed by fusing the

Table 1 Measurement of BSt Transcription in *E. coli* containing pSK102, pSK106 or pSK101

		pSK102	pSK106	pSK101
BSt expression		<0.01%	>2%	>20%
Pulse-labeled mRNA -	cpm*	3130	4690	15,630
	relative	0.2	0.3	1

*Normalized by taking plasmid copy number into consideration.

HaeII-BglI fragment carrying BSt cDNA to the SmaI site in pSK3 (21). As shown in Fig. 2, plasmid pSK106 carries the BSt sequence with the first codon for alanine changed to that for phenylalanine and pSK101 has the BSt sequence fused to 16 codons from a fusion of trpL to trpE (22). The levels of BSt accumulated in *E. coli* K12 containing pSK106 and pSK101 were 2-5% and 20-30%, respectively, of total cellular protein. Analysis of BSt expression by SDS-PAGE is shown in Fig. 3. The high level expression from pSK101 is evident from the prominent band indicated by the arrow. These results indicate that changes at the beginning of the BSt cDNA can affect expression, i.e., modification of the first codon increases expression by more than 200-fold (from less than 0.01% in pSK102 to greater than 2% in pSK106) and fusion of the cDNA to an *E. coli* sequence increases expression by more than 2,000-fold (from less than 0.01% in pSK102 to greater than 20% in pSK101).

Poor BSt Expression Is Not Due To Low Levels or Degradation of mRNA

To determine if the lack of BSt expression from the unmodified cDNA is due to low levels of mRNA or the degradation of mRNA at the 5' end, measurements of BSt-specific mRNA were carried out. To determine the level of BSt-mRNA, cells containing pSK102, pSK106 or pSK101 induced for expression were pulse-labeled with [5-^{3}H]uridine and the labeled RNA was hybridized with a DNA probe coding for 85% of the BSt sequence. To calculate the relative level of BSt-mRNA from these 3 plasmids, plasmid copy number was taken into consideration. The determined average plasmid copy number for pSK102, pSK106 and pSK101 was 57, 48 and 78, respectively. Table 1 shows that there was only a 5 fold difference in BSt specific pulse-labeled message between pSK102 and pSK101 and about 3 fold difference between pSK106 and pSK101. These results show that although the expression of BSt in pSK102 and pSK106 is about 2,000-fold and 10-fold, respectively, lower than that in pSK101, the level of pulse labeled BSt-mRNA differs only 3- to 5-fold in these strains.

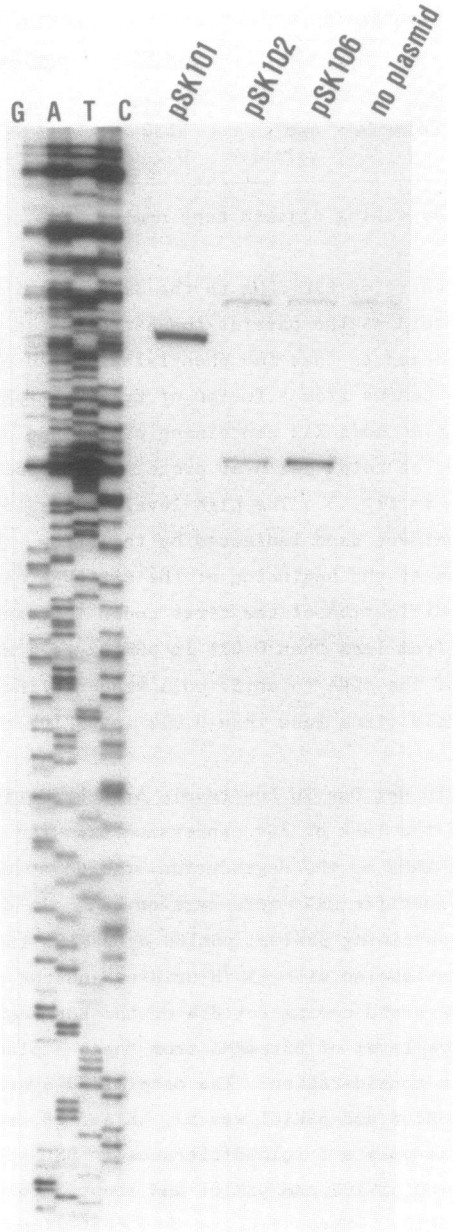

Fig. 4 Autoradiograph of a dideoxy-sequencing ladder and the primer extension products from cells containing pSK101, pSK102, pSK106 and no plasmid.

Using the DNA probe coding for 85% of BSt, Northern analysis was carried out to determine the size and steady state level of the BSt-mRNA with RNA isolated from cells containing pSK102, pSK106 and pSK101 incubated in the presence of rifampicin. To serve as a relative standard, a DNA fragment containing β-lactamase sequence was also used as a probe in the Northern analysis. The results indicate that the size of the BSt-specific message in all 3 strains is about 1 kb. The size of the β-lactamase mRNA is at the length expected (16).

Although the levels of mRNA were similar, it was possible that the messages from pSK102 and pSK106 were not translated due to heterogeneity or degraded 5' ends. To determine the integrity of the 5'-end of the BSt mRNA, primer extension of message isolated from cells containing pSK102, pSK106 and pSK101 was performed with a [^{32}P]-labeled oligonucleotide complimentary to codons 63-69 of BSt. A dideoxy-sequencing ladder of these plasmids using the same oligonucleotide as a primer served as a marker for the 5'-end of the mRNA. Fig. 4 shows an autoradiograph of a 6% denaturing polyacrylamide gel of a dideoxy-sequencing ladder and the primer extension product from pSK101, pSK102 and pSK106. Several things should be noted in this figure. 1) The 5' end of the messages from all 3 plasmids are relatively intact since no lower molecular weight species appear on the gel. The larger band seen in pSK102 and pSK106 is host-derived since this band is also observed in cells without a plasmid. 2) The start sites of the messages from all three plasmids are in the spot expected from the trp promoter (23). The primed RNA from pSK101 is larger than that from pSK102 and pSK106, as expected since in pSK101 the BSt gene is fused to 16 codons from trpLE (see Fig. 2). 3) The relative amount of the primed products corresponds to that observed by pulse labeling, i.e., pSK101 > pSK106 > pSK102.

These results demonstrate that in pSK102, where BSt expression is minimal, the 5' ends of the BSt-specific message are intact and the BSt-mRNA level is not much lower than that in pSK101 which expressed BSt to high levels. Taken together, analysis of the BSt-specific mRNA suggests that transcription is not a major problem in the lack of expression. Thus it is likely that poor expression is due to deficiency in translation. Since a simple change in the first codon of BSt (pSK106) results in more than a 200-fold increased expression over the unmodified cDNA sequence (pSK102), it is reasonable to suggest that this deficiency may be caused by the sequence at the beginning of the BSt gene. To test this possibility, additional

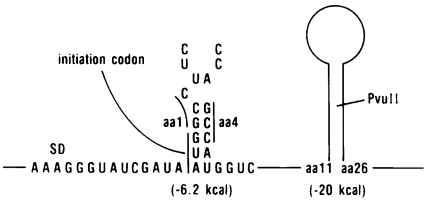

Fig. 5 Putative secondary structure in mRNA at the beginning of the BSt coding region in plasmid pSK102 carrying the unmodified cDNA.

modifications at the beginning of the BSt gene were made and their effect on translation initiation analyzed.

mRNA Secondary Structure Around Codons 12-25 Does Not Affect Expression

In examining the sequence of BSt-mRNA derived from the unmodified cDNA (pSK102), two putative secondary structures at the beginning of the coding region were observed (Fig. 5). The first stem-loop is weak with a free energy of -6.2 kcal/mol (based on Queen and Korn, reference 18). The second stem-loop can be formed using codons for amino acid residues 12-25 and, if formed, would be predicted to be -20 kcal/mol. The high level expression observed with pSK101 (see Fig. 3) suggests that this second stem-loop does not affect translation. However, it is possible that the position of this stem-loop may be a factor. In pSK101, this stem-loop is 78 bases from the initiating AUG while in BSt cDNA (pSK102), the distance is 33 bases. To test if this stem-loop interferes with translation when located 33 bases

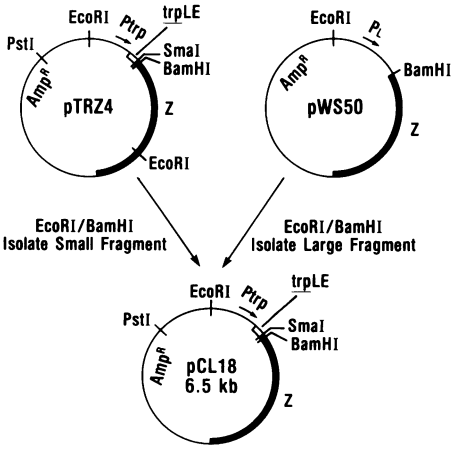

Fig. 6 Construction of the vector pCL18 for use in gene fusions to *lacZ*.

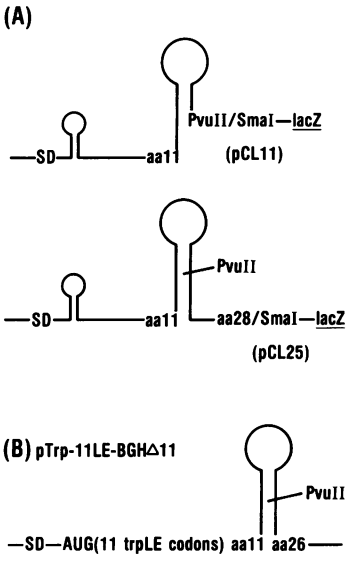

Fig. 7 Secondary structure in pCL11, pCL25 and pTrp-11LE-BGHΔ11.

from the initiation codon, two types of experiments were carried out. One was to fuse the BSt sequence both within and downstream from the stem-loop with a truncated lacZ, such that translation of lacZ is dependent upon initiation at the BSt AUG. The second experiment was to replace the first 11 codons for BSt with 11 codons from trpLE fusion.

To provide a vector for constructing the fusions to lacZ, plasmid pCL18 was made. As shown in Fig. 6, the small EcoRI-BamHI fragment containing the trp promoter was isolated from pTRZ4 (obtained from Leonard Post, The Upjohn Company) and used to replace the small EcoRI-BamHI region in pWS50 (10). The resulting pBR-background vector pCL18 has a lacZ sequence deleted for the first 8 codons. This truncated lacZ is downstream from the trp promoter and 16 codons of a trpLE fusion and is not in frame with the trpLE codons. To make the fusion of the beginning of BSt sequence without the second stem-loop to lacZ, the small EcoRI-SmaI region in pCL18 was replaced with the small EcoRI-PvuII fragment carrying the trp promoter and the first 23 BSt codons isolated from pSK102. The resulting plasmid pCL11 (Fig. 7A) does not have the second stem-loop and gives 658 units of β-galactosidase activity. To make the fusion of the beginning of BSt containing the second stem-loop to lacZ, pSK102 was digested with SmaI (located at codon 133 of the BSt gene), treated with exonuclease Bal31, digested with EcoRI, and the small

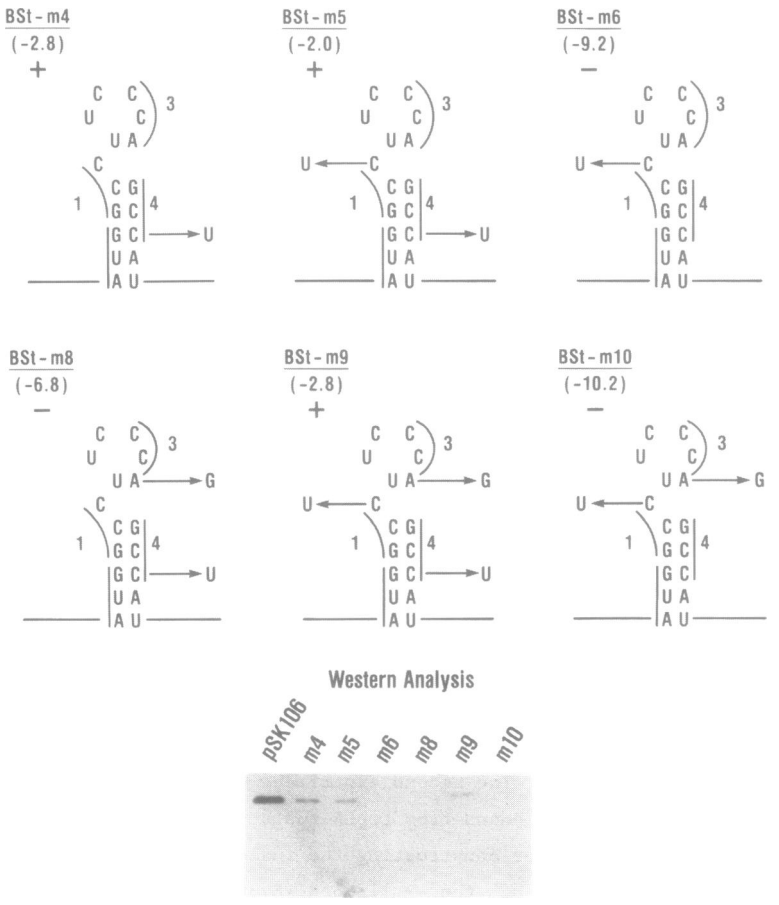

Fig. 8 Alterations in BSt codon 1, 3 and 4 to manipulate the stability of the stem-loop structure. The changes made are illustrated in the 6 sketches. The free energy (in kcal) of the stem-loops are shown in parenthesis and BSt expression is indicated by + or -. The stem-loop in pSK102 (no modifications) has a free energy of -6.2 kcal and the expression is - and undetectable by Western hybridization. The level of expression is shown in the Western analysis.

fragments isolated and used to replace the small EcoRI-SmaI region in pCL18. One of the resulting plasmids, pCL25, has the first 28 codons of BSt fused to lacZ (Fig. 7A). This plasmid retains the second stem-loop and gives 839 units of β-galactosidase activity. These results demonstrate that the level of expression is similar with or without the presence of the second stem-loop structure in the beginning of the mRNA, indicating that this stem-loop does not interfere with translation.

To replace the first 11 BSt codons with 11 codons from trpLE, oligonucleotides containing the trpLE codons were inserted upstream of the truncated BSt sequence in pTrp-BGHm1b (see Fig. 1). The resulting vector pTrp-11LE-BGHΔ11 can direct the expression of the LE-BGH to greater than 20% of total cell protein. In this replacement, the strong second stem-loop is present at the same location as in pSK102 (see Fig. 5 and Fig. 7B) where the expression level is at 0.01%. This result agrees with the findings from the above lacZ fusions, i.e., the presence of the strong secondary structure does not block translation. It also supports the idea that the sequence within the beginning of the BSt cDNA (within the first 11 codons) is not optimal for efficient translation initiation.

mRNA Secondary Structure Around Initiation Codon Interferes with Expression

As shown in Fig. 5, there is a putative weak stem-loop at the beginning of the BSt-mRNA from the unmodified cDNA sequence which expresses BSt at insignificant levels. This stem-loop includes the initiation codon AUG and BSt codons 1 to 5, making the AUG inaccessible. To test if this stem-loop hinders translation initiation, modifications without changing the amino acid residues were made in codon 4 or codons 1 and 4 to remove the stem-loop (see Fig. 1 for procedures). A further alteration was made in codon 3 to reinstate the stem-loop. These changes and their effect on expression are shown in Fig. 8. As can be seen in Fig. 8, changes in codon 4 for alanine from GCC to GCU (BSt-m4) reduces the stem-loop of -6.2 kcal in the unmodified cDNA (pSK102) to -2.8 kcal and changes in both codons 1 and 4 for alanine from GCC to GCU (BSt-m5) reduce the stem-loop to -2.0 kcal. Both BSt-m4 and BSt-m5 produce BSt while pSK102 does not. Changing codon 1 for alanine from GCC to GCU (BSt-m6) does not reduce the stability of the stem-loop (the free energy is increased to -9.2 kcal) and does not lead to BSt expression. If disrupting the stem-loop structure by the changes made in BSt-m4 and BSt-m5 is responsible for increasing expression, restoring the stem-loop while keeping the BSt-m4 change should abolish expression. In BSt-m8, the stem-loop removed by the change made in BSt-m4 is restored to -6.8 kcal by changing the codon 3 for proline from CCA to CCG. This reinstatement of the stem-loop results in no BSt expression. The fact that the effect on expression from the alteration in codon 3 is due to the increased stem-loop stability and not due to the change per se is supported by the result observed with BSt-m9. BSt-m9 has the same change in codon 3 as BSt-m8, but since the stem-loop remains disrupted the expression is not prohibited. BSt-m10 has changes in both codons 1 and 3 resulting in a more

stable stem-loop of -10.2 kcal and no expression is observed. The level of expression in m4, m5 and m9 is about 0.5-1% of total cell protein, determined by comparison to pSK106 which gives about 2-5% expression (see Western analysis in Fig. 8).

DISCUSSION

Expression of the native BSt cDNA in *E. coli* is poor due to inefficient translation initiation. Although it has been suggested that the nature of the block may be due to secondary structure in the mRNA around the Shine-Dalgarno sequence and/or in the beginning of the gene, such a model has not been thoroughly analyzed. Here we report the analysis of several modifications in the BSt gene downstream from the initiation codon to ascertain the level and integrity of the mRNA and to test if any secondary structures in the mRNA are responsible for blocking expression. We conclude that a stem-loop structure in the mRNA with a free energy of -6.2 kcal/mole involving the initiation codon and the first 5 codons in the BSt gene has a strong negative effect on expression. However, elimination of this structure with different single or multiple substitutions result in varied levels of expression. Thus, in addition to the secondary structure, information (or lack of it) encoded in the first few codons also plays a role in determining the level of translation of this gene. Furthermore, the presence of a strong stem-loop structure (-20 kcal/mol) beginning at the 12th codon from the start does not appear to interfere with expression.

Although the stem-loop involving the initiation codon AUG (see Fig. 5) is relatively weak, the changes made within this stem-loop (m4, m5, m6, m8, m9 and m10; see Fig. 8) clearly demonstrate that this secondary structure is responsible for lowering expression. There are several examples of such structures around initiation codons that hinder translation initiation (24-31). It is most likely that these structures interfere directly with the ability of the mRNA to interact with the ribosome-fmet-tRNA complexes. Although it is possible that the single base change in BSt-m4 (GCC to GCT) increases expression by changing the interactions between the mRNA and initiation complex in a manner independent of secondary structure, the following observations argue against this possibility. In m4, m5, m6, m8, m9 and m10, the changes were made only in wobble positions so as to maintain the same amino acid residues in the protein. Although the difference in the level of expression among these changes could be due to the specific tRNA used for that codon, since all the changes that increase expression can be

suppressed by restoring the stem-loop with a change at another position, we feel that these changes per se do not significantly alter the interactions between the mRNA and the initiation complex by a mechanism independent of secondary structure.

If the difference in expression between pSK102 (unmodified cDNA; <0.01%) and BSt-m4 (0.5-1%) is due to secondary structure around the AUG, what is responsible for the differences observed between BSt-m4 (0.5-1%), pSK106 (2-5%) and pSK101 (20-30%). The slight increase in expression between BSt-m4 ($alanine_4$ from GCC to GCU) and pSK106 ($alanine_1$ GCC to phenylalanine UUC) could be explained by the possibility of G:U base pairing in BSt-m4 mRNA that would maintain some of the inhibitory structure (see Fig. 5). We have no data to support this model, but it can be tested directly by altering the GCC for $alanine_4$ to GCA or GCG. We have ruled out the possibility that the difference in expression between pSK106 (2-5%) and pSK101 (20-30%) is due to the presence of a putative stem-loop structure extending from codon 12 to 25 shown in Fig. 5. Since this structure is 33 bases from the initiating AUG in pSK106 and 78 bases in pSK101, its location with respect to the start codon may play some role. However, the fact that the two BSt-lacZ fusions pCL11 and pCL25 (lacZ fused to BSt within and downstream from the stem-loop) yielded similar levels of β-galactosidase, and that pTrp-11LE-BGHΔ11 (the stem-loop is 33 bases from the initiating AUG) produced the same amount of BSt as pSK101 suggests that this structure does not significantly influence expression. Although we have examined the role of the two obvious stem-loop structures at the beginning of the gene, we cannot rule out that some other structure is responsible for the difference between pSK106 and pSK101. Another possibility for the difference in BSt levels between pSK106 and pSK101 could be that the fusion protein from pSK101 is more stable than the protein molecules synthesized from pSK106. Pulse-chase labeling experiments, however, show that the fusion and nonfusion proteins are equally stable under the conditions used in these experiments (N. Watson, personal communication). Therefore there must be some element residing in the sequence downstream from the AUG in pSK101 or pTrp-11LE-BGHΔ11 that can enhance expression. Alternatively, some component downstream from AUG in pSK106 and BSt-m4 is acting in a negative fashion or is lacking the information for optimal initiation. This sequence may act either by itself or in combination with the particular sequence upstream of the AUG that we provide in the vector. Interaction of this BSt

gene sequence with upstream region in translation initiation is supported by the observation that placement of the BSt sequence in BSt-m4 behind a ribosome binding site rich in A and T bases greatly increase expression (4). The enhancing effect of this AT-rich ribosome binding site on BSt-m4 expression does not appear to be caused by obvious mRNA secondary structure upstream of the initiation codon. Distinguishing the involvement of such interactions in the initiation steps will require further mutational and biochemical analysis.

ACKNOWLEDGEMENT

We thank Drs. Don Court and John Mott for helpful discussions and N. Watson for providing unpublished data.

*To whom correspondence should be addressed

REFERENCES

1. Seeburg, P.H., Sias, S., Adelman, J., deBoer, H.A., Hayflick, J., Jhurani, P., Goeddel, D.V., and Heyneker, H.L. (1983) DNA 2, 37-45.
2. Schoner, B.E., Hsiung, H.M., Belagaje, R.M., Mayne, N.G., and Schoner, R.G. (1984) Proc. Natl. Acad. Sci. USA 81, 5403-5407.
3. George, H.J., L'Italien, J.J., Pilacinski, W.P., Glassman, D.L., and Krzyzek, R.A. (1985) DNA 4, 273-281.
4. Olsen, M.K., Rockenbach, S.K., Curry, K.A., and Tomich, C-S.C. (1988) J. Biotechnology, In Press.
5. Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982) Molecular Cloning. Cold Spring Harbor Laboratory, New York.
6. Wallace, R.B., Johnson, M.J., Suggs, S.V., Miyoshi, K-I., Bhatt, R., and Itakura, K. (1981) Gene 16, 21-26.
7. Towbin, H., Staehelin, T., and Gordon, J. (1979) Proc. Natl. Acad. Sci. USA 76, 4350-4354.
8. Morse, L.S., Pereira, L., Roizman, B., and Schaffer, P.A. (1978) J. Virol. 26, 389-410.
9. Friedman, D.I., Jolly, C.T., and Mural, R.J. (1973) Virol. 51, 216-226.
10. Sisk, W.P., Chirikjian, J.G., Lautenberger, J., Jorcyk, C., Papas, T.S., Berman, M.L., Zagursky, R., and Court, D.L. (1986) Gene 48, 183-193.
11. Miller, J.H. (1972) Experiments in Molecular Genetics, p. 48. Cold Spring Harbor Laboratory, New York.
12. Court, D., DeCrombrugghe, B., Adhya, S., and Gottesman, M. (1980) J. Mol. Biol. 138, 731-743.
13. Bogosian, G., Somerville, R.L., Nishi, K., Kano, Y., and Imamoto, F. (1984) Mol. Gen. Gent. 193, 244-250.
14. Baughman, G., and Nomura, M. (1983) Cell 34, 979-988.
15. Messing, J. (1983) Methods in Enzymology 101, 20-78.
16. VonGabain, A., Belasco, J.G., Schottel, J.L., Chang, A.C.Y., and Cohen, S.N. (1983) Proc. Natl. Acad. Sci. USA 80, 653-657.

17. Polites, H.G., and Marotti, K.R. (1986) Biotechniques 4, 514-520.
18. Queen, C., and Korn, L.J. (1984) Nucleic Acids Res. 12, 581-599.
19. Tinoco, I. Jr., and Nussinov, R. (1982) Nucleic Acids Res. 10, 341-349.
20. Salser, W. (1977) Cold Spring Harbor Symposium on Quantitative Biology 42: 985-1002.
21. Kaytes, P.S., Theriault, N.Y., Poorman, R.A., Murakami, K., and Tomich, C-S.C. (1986) J. Biotechnology 4, 205-218.
22. Miozzari, G.F., and Yanofsky, C. (1978) J. Bacteriol. 133, 1457-1466.
23. Bennett, G.M., Schweingruber, M.E., Brown, K.D., Squires, C., and Yanofsky, C. (1978) J. Mol. Biol. 121, 113-137.
24. Hall, M.N., Gabay, J., Debarbouille, M., and Schwartz, M. (1982) Nature 295, 616-618.
25. Tessier, L-H., Sondermeyer, P., Faure, T., Dreyer, D., Benavente, A., Villeval, D., Courtney, M., and Lecocq, J-P. (1984) Nucleic Acids Res. 12, 7663-7675.
26. Wood, C.R., Boss, M.A., Patel, T.P., and Spencer-Emtage, J. (1984) Nucleic Acids Res. 12, 3937-3950.
27. Munson, L.M., Stormo, G.D., Niece, R.L., and Reznikoff, W.S. (1984) J. Mol. Biol. 177, 663-683.
28. Coleman, J., Inouye, M., and Nakamura, K. (1985) J. Mol. Biol. 181, 139-143.
29. Stanssens, P., Remaut, E., and Fiers, W. (1985) Gene 36, 211-223.
30. Looman, A.C., Bodlaender, J., deGruyter, M., Vogelaar, A., and vanKnippenberg, P.H. (1986) Nucleic Acids Res. 14, 5481-5497.
31. Gold, L. (1988) In Annual Review of Biochemistry, (C.C. Richardson, P.D. Boyer, I.B. Dawid and A. Meister, eds.) Annual Reviews Inc., Palo Alto, Vol. 57, pp. 199-234.