Volume 17 Number 5 1989 Nucleic Acids Research

Use of synthetic ribosome binding site for overproduction of the 5B protein of insertion sequence IS*5*

Jeffrey M.Chernak*+ and Hamilton O.Smith

Department of Molecular Biology and Genetics, The Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA

Received November 21, 1988; Revised and Accepted February 6, 1989

ABSTRACT

Insertion sequence IS5 is a bacterial transposable element which contains three open reading frames designated 5A, 5B and 5C. Although there was no detectable expression from the 5B open reading frame when it was preceded by the native promoter and ribosome binding site or by a tac promoter and the native ribosome binding site, we have overproduced a 5B protein both in vitro and in Escherichia coli cells by using a tac promoter and a specially-designed synthetic ribosome binding site. ß-galactosidase fusion studies suggested that the synthetic binding site is at least 150-fold more efficient than the native binding site. The 5B protein amounted to 80-85% of the total protein made in vitro and 20-25% of the total protein pulse-labelled in whole cells. It is stable in vitro but rapidly degraded in vivo. Thus expression of the 5B gene appears to be limited by both poor translation initiation and protein degradation.

INTRODUCTION

Insertion sequences (IS) are small procaryotic transposable elements (0.75-1.5 kilobases) which generally contain one long open reading frame (ORF; for reviews, see 1,2). Gene products have been detected from the long ORFs of IS1, IS4, IS5 and IS50 (3-7). There is often at least one other ORF which is shorter than the first and is encoded on the opposite DNA strand (2,4,8). Although there is some evidence that these smaller ORFs may also produce specific proteins (4,8), studies on their expression and regulation are limited.

The 1195-base pair (bp) insertion sequence IS5 (Fig. 1) has a compact genetic structure which includes three ORFs (9-11) and associated promoter and terminator signals (4,8,10). The longest ORF has been shown to encode a 37-kilodalton (kDa) 5A protein in vitro and in Escherichia coli minicells or maxicells (3,4), and the smaller ORFs have been reported to encode a 12.3-kDa 5B pro-

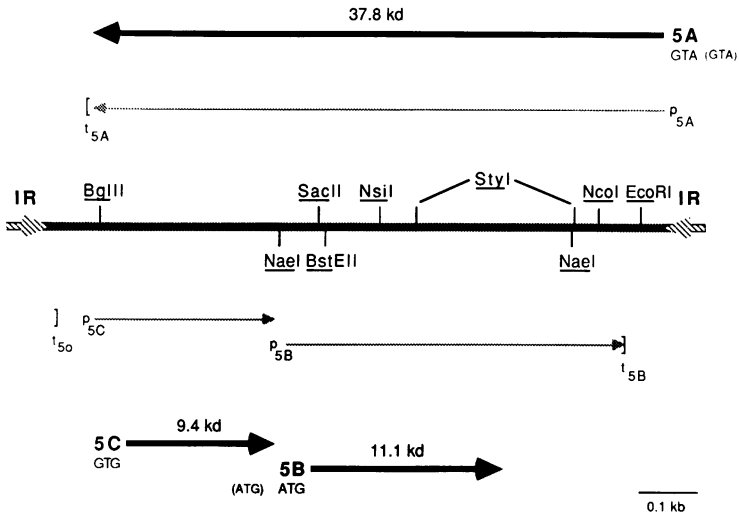


FIG. 1. Structure of insertion sequence IS5. The striped arrows represent terminal inverted repeats (IR) which are identical at 15 out of 16 positions. Promoters (p_{5A}, p_{5B}, p_{5C}) and terminators (t_{5A}, t_{5B}, t_{5o}) are shown at their appropriate locations either upstream or downstream of the three putative RNA transcripts, which are designated by thin arrows. The polarity of the terminators is illustrated by brackets. The three open reading frames (5A, 5B, 5C) are indicated by thick solid arrows, with the predicted size of the corresponding gene products given in kilodaltons (kd). In each case, the probable translational start codon is shown preceding the arrow; less likely upstream starts are indicated in parentheses. The thin bar at the lower right gives the approximate scale of the diagram, in kilobases (kb).

tein and a 9.4-kDa 5C protein in minicells (4,8). The observed molecular weight of the 5B protein differed somewhat from the predicted value of 11.1 kDa. In previous work, we have been unable to confirm the synthesis of a 12.3-kDa species from an intact copy of IS5, even when the 5B ORF was located downstream of a strong tac promoter (3,12).

This study demonstrates the in vitro and in vivo synthesis and overproduction of the 5B protein using a specially-designed synthetic ribosome binding site (SRBS) in conjunction with a tac promoter and a subcloned copy of the 5B ORF. Although this protein is stable in vitro, it is rapidly degraded in vivo. Thus, expression of the 5B gene appears to be limited by both poor translation initiation and protein degradation. The 5B protein

reported here has an apparent molecular weight of 8.4 kDa on denaturing sodium dodecyl sulfate(SDS)-polyacrylamide gels, in contrast to both the predicted value of 11.1 kDa and the previously-reported value of 12.3 kDa. This discrepancy is discussed below.

MATERIALS AND METHODS

Bacteria and plasmids

E. coli strains HB101 *hsdS*20 ($r_B^-m_B^-$) *recA*13 (13), RB791 *laci*^{q} L8 (14), JM109 *hsdR*17 ($r_K^-m_K^+$) *recA*1 Δ(*lac*-*pro*AB)/F' (*pro*AB, *laci*^{q}ZΔM15) (15), MM294 *hsdR*17 ($r_K^-m_K^+$) (16) and W3110 (17) have been described. Growth media included L (18), A and M9 media (19). Antibiotic concentrations were 100 μg/ml ampicillin and 20 μg/ml tetracycline.

Plasmids pBR322 (20), ptacI1 (14), and pMLB1034 (18) have been previously described; pSK10 (21) carries a promoterless galactokinase gene downstream of a *tac* promoter and contains a *laci*^{q} gene cloned into the *Eco*RI site of pDR540 (22). The plasmids pDI15 and pDI40 contain IS*5* at the same site but in opposite orientations within the pBR322 derivative pDI14 (11), and pJC100 contains IS*5* inserted downstream of the pBR322 *tet* promoter (3). Plasmids pJC200, pJC201, pJC251, pJC281, pJC287 and pJC57 are described below.

Preparation, analysis, and manipulation of DNA

Plasmids and DNA fragments were purified, analyzed, and gel-purified by standard methods (23). Restriction enzymes, T4 DNA ligase, T4 polynucleotide kinase, T4 DNA polymerase, Klenow fragment of *E. coli* DNA polymerase I, and bacterial alkaline phosphatase were obtained from New England Biolabs or IBI, and used as recommended. Ligations and other recombinant DNA techniques were carried out as previously described (23).

Construction of pJC200, pJC201, and pJC251

The 509-bp *Nae*I fragment of pDI40 containing the 5B ORF and a portion of its promoter was ligated to *Eco*RV-digested pBR322 and ampicillin-resistant, tetracycline-sensitive HB101 transformants were tested for the presence and orientation of an insert in the resident plasmid. A plasmid containing the 5B ORF downstream of the pBR322 *tet* promoter was designated pJC201 and a plasmid con-

taining the insert in the opposite orientation was designated pJC200.

Plasmid pSK10 was digested with BamHI, blunt-ended with a T4 polymerase fill-in reaction, and ligated to the 509-bp NaeI fragment from pDI40. Ampicillin-resistant HB101 transformants were screened for the presence and orientation of an insert in their resident plasmids, and a plasmid containing the 5B ORF downstream of the tac promoter was designated pJC251.

Construction of pJC281

The top and bottom strands of the SRBS (see Fig. 3) were synthesized, purified and annealed. Plasmid pJC100 was digested first with BamHI and EcoRI, and then with SacII. The 570-bp SacII-EcoRI fragment containing the 3'-end of the 5B ORF was ligated to the double-stranded EcoRI-SacII SRBS containing the 5' end of the 5B ORF in the presence of EcoRI, and the 600-bp fragment containing the SRBS and intact 5B ORF was gel-purified, phosphorylated and ligated to dephosphorylated, EcoRI-digested ptac11 vector. Ampicillin-resistant RB791 $laci^q$ transformants were screened for the presence of the correct plasmid by colony hybridization (23,24) with a radiolabelled SRBS probe. The resident plasmid of one of the positive clones was shown by DNA sequence analysis (25) to contain the expected SRBS sequences and intact 5B ORF downstream of the tac promoter (12); this plasmid was designated pJC281.

Construction of pJC287 and pJC57

BstEII-digested pJC281 was blunt-ended by a fill-in reaction with Klenow fragment of DNA polymerase I and ligated to SmaI-linearized pMLB1034. Ampicillin-resistant JM109/F'$laci^q$ transformants were selected on minimal agar plates containing either ampicillin and Xgal (5-bromo-4-chloro-3-indolyl-ß-D-galactoside; Sigma) or ampicillin, Xgal and IPTG (isopropyl-ß-D-thio-galactoside; Boehringer Mannheim). Colonies with a blue phenotype were analyzed by restriction mapping of their resident plasmids, and a plasmid containing the promoterless ß-galactosidase gene downstream of the tac promoter and SRBS was designated pJC287.

Plasmid pBR322 was digested with HincII and ligated to SmaI-linearized pMLB1034. Ampicillin-resistant JM109/F'$laci^q$ transformants were selected and analyzed as above, and a plasmid con-

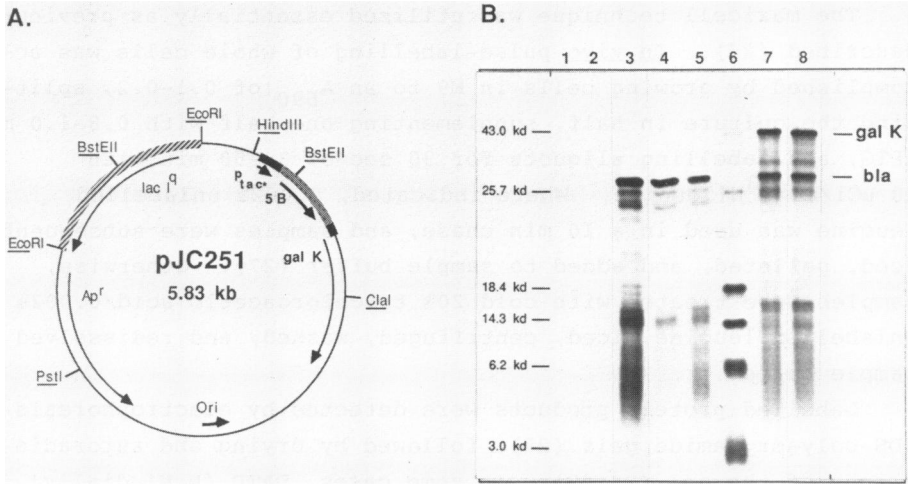


FIG. 2 (A). Structure of the plasmid pJC251. The NaeI fragment of pDI40 containing the 5B ORF and a portion of its promoter was cloned into the BamHI site of pSK10 to yield pJC251, in which the 5B ORF is located between the strong tac promoter (p_{tac*}) and the promoterless galactokinase gene (gal K). IS5 sequences are designated by a thick shaded arc, an EcoRI fragment containing a $lacI^q$ gene is indicated by a thick striped arc, and the ampicillin resistance gene (Ap^r) from pBR322 is also shown. (B). In vitro synthesis of galactokinase from pJC251 demonstrates transcription of the 5B ORF. Proteins were synthesized by in vitro transcription-translation using 25 µCi of [^{35}S]methionine and 1.0-1.2 µg of substrate DNAs. Where indicated, reactions were supplemented with 1 mM IPTG. Proteins were analyzed on a 17% SDS-polyacrylamide gel. Lanes: 1, no DNA, no S30 extract; 2, no DNA; 3, pBR322; 4, pJC200; 5, pJC201; 6, molecular weight markers; 7, pJC251 + IPTG; 8, pSK10 + IPTG. All lanes are from the same gel. Positions of bands corresponding to galactokinase (gal K) and ß-lactamase (bla) are indicated on the right and molecular weights of protein standards are given in kilodaltons (kd) on the left.

taining a fusion of the tetracycline resistance gene and the ß-galactosidase gene was designated pJC57.

Labelling and analysis of plasmid-encoded proteins

In vitro transcription-translation reaction components were purchased from the Codon Company (Mundelein, IL) and used essentially as recommended. Labelling reactions utilized [^{35}S]methionine (1100 Ci/mmol, 10-50 mCi/ml; NEN) or [^{3}H]leucine (120-160 Ci/mmol, 1-5 mCi/ml; NEN).

The maxicell technique was utilized essentially as previously described (26). In vivo pulse-labelling of whole cells was accomplished by growing cells in M9 to an A_{590} of 0.1-0.2, splitting the culture in half, supplementing one half with 0.8-1.0 mM IPTG, and labelling aliquots for 30 sec or 30-50 min with 10 μCi/ml [^{3}H]leucine. Where indicated, 0.002% unlabelled leucine was used in a 10 min chase, and samples were subsequently iced, pelleted, and added to sample buffer (27). Otherwise, samples were treated with cold 20% trichloroacetic acid/0.002% unlabelled leucine, iced, centrifuged, washed, and redissolved in sample buffer.

Labelled protein products were detected by electrophoresis on SDS-polyacrylamide gels (27), followed by drying and autoradiography of the gel (28,29). In some cases, DATD (N,N'-diallyltartardiamide; Aldrich) was used as the cross-linker rather than bis-acrylamide. Radiolabelled protein standards (BRL) were as follows, with their molecular weights given in kilodaltons: ovalbumin (43.0), α-chymotrypsinogen (25.7), ß-lactoglobulin (18.4), lysozyme (14.3), bovine trypsin inhibitor (6.2), insulin (3.0).

Assay of ß-galactosidase activity

Relative levels of ß-galactosidase activity were measured essentially as described by Miller (19).

RESULTS

Evidence for transcription of the 5B ORF

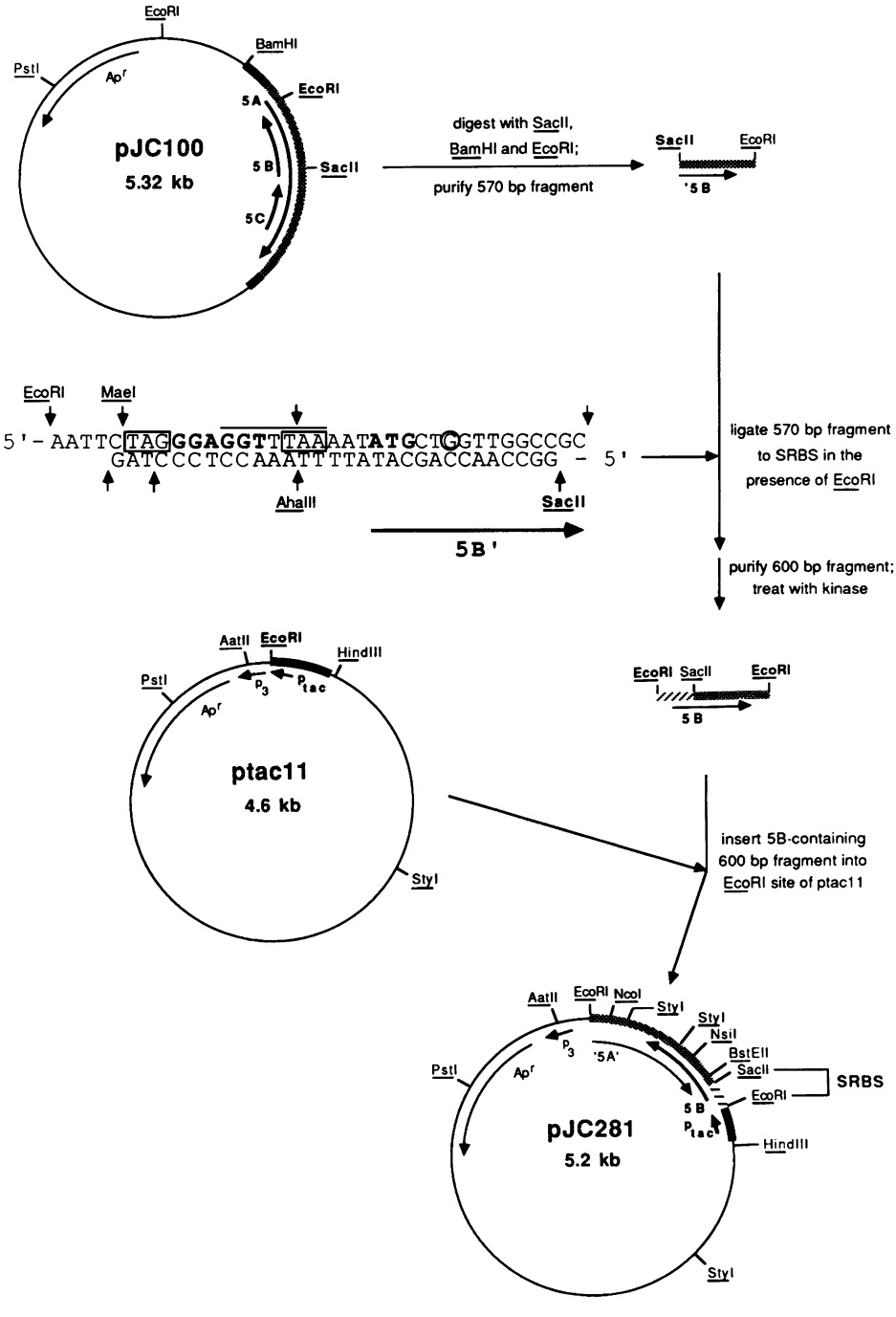
The plasmids pJC200 and pJC201 contain the 5B ORF and both of its possible ATG start codons subcloned in opposite orientations downstream of the pBR322 tet promoter. When these plasmids were used to direct the synthesis of proteins in a cell-free transcription-translation system (30), there was no convincing evidence for production of a protein of the expected size (11.1 kDa) from the 5B ORF (Fig. 2B). Plasmid pJC251, which contains the 5B ORF downstream of a strong tac promoter (see Fig. 2A), also failed to yield detectable levels of a product specific to this ORF. These results are consistent with those obtained earlier using an intact copy of IS5 (3,12). However, pJC251 contains a galactokinase gene downstream of the 5B ORF, and it produced a 43-kDa protein with a mobility corresponding closely

with that expected for galactokinase (31). Production of this protein was at the same level as that seen with the control plasmid pSK10. Thus it appears that transcription initiated at the tac promoter was proceeding through the 5B ORF and into the promoterless galactokinase gene but was not resulting in detectable levels of a 5B protein.

Use of the maxicell technique confirmed production of the 43-kDa protein in E. coli cells containing pJC251, but again there was no evidence for an 11.1-kDa 5B gene product (data not shown). Since the plasmid carries a $lacI^q$ gene (see Fig. 2A), synthesis of the 43-kDa species was induced in the presence of IPTG, supporting the conclusion that its production resulted from transcription from the tac promoter.

Replacement of the 5B translation initiation region with an efficient synthetic ribosome binding site

Since transcription of the 5B ORF does not result in appreciable levels of a 5B protein, it appears that either the 5B transcript is not readily translated or the 5B gene product is rapidly degraded. Examination of the IS5 sequences which might serve as the 5B ribosome binding site revealed several features which suggest that they are inefficient in directing translation initiation (see Discussion). Therefore, these sequences were replaced with a synthetic ribosome binding site (SRBS) which optimally incorporates those features which might contribute to efficient translation initiation (Fig. 3). The SRBS specifies a Shine-Dalgarno sequence with a good stretch of polypurines, good complementarity to the 16S rRNA (6 out of 6 important positions), and appropriate spacing from the ATG start (32,33); a heptanucleotide sequence of the form RRTTTRR (where R stands for a purine; see 34); a spacer region composed entirely of A's and T's (32,33,35) and preferred nucleotides at the -3, -2 and -1 positions (32,33,36); and a lack of any significant secondary structure which might interfere with utilization of the Shine-Dalgarno sequence or ATG start (37,38). In addition, translational readthrough from any upstream ORF is prevented by a TAA stop codon in the spacer region, while translation in the other two reading frames can be terminated by a TAG codon 5' to the Shine-Dalgarno of the SRBS or a TAA codon in the vector sequences


EcoRI
BamHI
PstI
Ap^r
5A
EcoRI
pJC100
5.32 kb
5B
SacII
5C
digest with SacII,
BamHI and EcoRI;
purify 570 bp fragment
SacII
EcoRI
'5B
EcoRI
MaeI
5'-AATTCTAGGGAGGTTTAAAATATGCTGGTTGGCCGC
GATCCCTCCAAATTTTATACGACCAACCGG - 5'
AhaIII
SacII
5B'
ligate 570 bp fragment
to SRBS in the
presence of EcoRI
purify 600 bp fragment;
treat with kinase
AatII
EcoRI
HindIII
PstI
P3
Ptac
Ap^r
ptac11
4.6 kb
StyI
EcoRI
SacII
EcoRI
5B
insert 5B-containing
600 bp fragment into
EcoRI site of ptac11
AatII
EcoRI
NcoI
StyI
StyI
NsiI
BstEII
SacII
SRBS
EcoRI
PstI
P3
'5A'
Ap^r
5B
Ptac
pJC281
5.2 kb
HindIII
StyI

of the final construct. The last 15 bp of the SRBS include a single change from the native sequence -- substitution of a G for a C to change the rarely utilized leucine codon CTC to the commonly utilized leucine codon CTG (39-41).

The efficiency of the SRBS in directing translation initiation was analyzed using plasmid pJC287, which contains the tac promoter, SRBS, and first 8 codons of the 5B ORF fused to a ß-galactosidase gene that is missing 8 amino-terminal codons. The amount of ß-galactosidase activity produced from this plasmid should be directly proportional to the amount of ß-galactosidase enzyme made (19), and is therefore a reflection of the translational efficiency of the SRBS. JM109/F' $lacI^q$ cells containing pJC287 produced approximately 30 units of activity in the absence of IPTG and 3,000 units in the presence of IPTG, while cells containing the control plasmid pMLB1034 produced $\leq$1 unit of activity and cells containing a plasmid bearing the ß-galactosidase gene fused to the tet regulatory signals (pJC57) produced $\leq$20 units (data not shown). This demonstrates both inducible production and efficient translation of messages encoded by an ORF located downstream of the tac-SRBS regulatory signals. The activity detected without induction was probably due to incomplete repression of the high copy number tac promoter by the low copy number $lacI^q$ allele. Controls done with $lacI^+lacZ^+$ MM294 or W3110 cells gave results consistent with those expected (19). Similar constructions designed to fuse the tac promoter and 5B

FIG. 3. Structure of plasmid pJC281: Subcloning the 5B ORF downstream of a strong tac promoter and synthetic ribosome binding site. The specially-designed synthetic ribosome binding site (SRBS), which is flanked by EcoRI and SacII restriction sites, contains a Shine-Dalgarno sequence and ATG start codon, shown in boldface type; a heptanucleotide sequence of the form RRTTTRR (R=purine), which is overscored; and upstream translational stop codons, which are enclosed in boxes. The circled G residue corresponds to a position at which the native 5B ORF contains a C (see Fig. 6). IS5 sequences are indicated by shaded arcs or boxes, the SRBS is designated by a striped box, and E. coli sequences containing the strong tac promoter (p_{tac}) are represented by a black arc. 5A, 5B and 5C indicate intact IS5 genes, while '5B, 5B', and '5A' designate genes that are truncated at the 5', 3', or 5' and 3' ends, respectively. The ß-lactamase promoter (p_3) and ampicillin resistance gene (Ap^r) are also shown. Details of the construction are described in Materials and Methods.

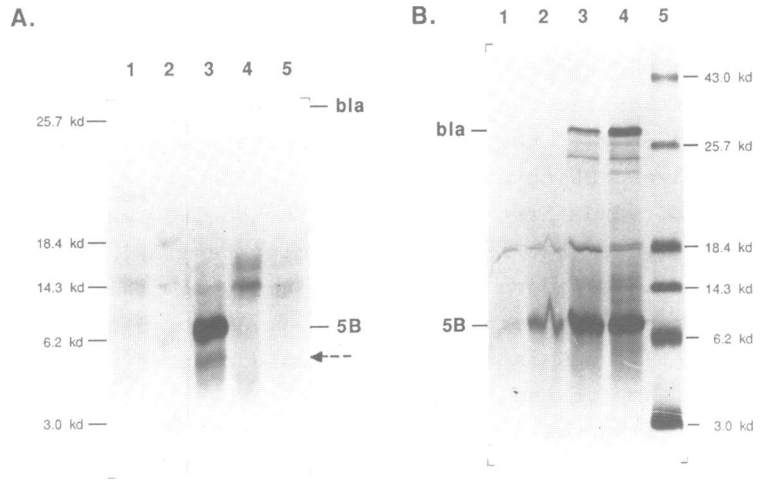


FIG. 4 (A). In vitro overproduction of the 5B protein from pJC281. Proteins were synthesized by in vitro transcription-translation using 50 µCi of [^{3}H]leucine and 0.4-0.8 µg of the indicated substrate DNAs in a 40 min reaction. Aliquots of labelled proteins containing 170-300 x 10^3 cpm were electrophoresed on a 20% SDS-polyacrylamide gel and the gel was autoradiographed for 4 hr. Lanes: 1, pDI15; 2, molecular weight markers; 3, pJC281; 4, ptac 11; 5, pJC251. All lanes are from the same gel. Positions of bands corresponding to ß-lactamase (bla), the 5B protein (5B), and an unidentified polypeptide of approximate molecular weight 5.5 kDa (<---) are indicated on one side of the figure, and molecular weights of protein standards are given on the other side.
(B). The 5B protein is stable in vitro. Proteins labelled with [^{35}S]methionine were synthesized from pJC281 (1.4 µg) by in vitro transcription-translation. Reaction mixtures were preincubated for 3 min before addition of the S30 extract, and there was no chase with unlabelled methionine following the labelling reaction. Labelled proteins containing 350-400 x 10^3 cpm were analyzed on a 20% SDS-polyacrylamide gel. Lanes: 1, 2 min reaction; 2, 5 min reaction; 3, 10 min reaction; 4, 30 min reaction; 5, molecular weight markers.

ribosome binding site from pJC251 to ß-galactosidase suggested that the SRBS is at least 150-fold stronger than the native 5B ribosome binding site. This data supports the hypothesis that the inability to detect a 5B gene product in previous experiments was due to inefficient translation initiation.

Overproduction of the 5B protein in vitro

The in vitro transcription-translation system was used to synthesize proteins from pJC281, which contains the 5B ORF down-

stream of the tac promoter and SRBS (see Fig. 3). As shown in Figure 4A, a protein with an apparent molecular weight of approximately 8.4 kDa was produced at a level of 80-85% of the total protein synthesized. This protein was not detected from either the control plasmid ptacll or from plasmids pDIl5 or pJC25l, which contain the 5B ribosome binding site and ORF downstream of the ß-lactamase promoter and 5B promoter or the tac promoter, respectively. This data strongly suggests that the 8.4-kDa protein is the product of the 5B ORF and that its production in such large amounts from pJC28l is due to the presence of the SRBS rather than the native ribosome binding site. Furthermore, it compares favorably with the prediction from the ß-galactosidase fusion studies that expression of a 5B protein would be at least l50-fold greater from pJC28l than from pJC25l. A 5.5-kDa species seen in Figure 4A was not consistently observed in vitro and did not appear to be made in vivo (Fig. 5). The results in Figure 4B indicate that the 8.4-kDa species is synthesized in detectable amounts within 2 min in vitro, and that it accumulates during the course of a 30 min reaction. Thus it appears that this product is relatively stable in vitro.

To test the hypothesis that the 8.4-kDa protein originates from the 5B ORF, pJC28l was linearized at sites both within and outside the ORF and then utilized in the in vitro transcription-translation system. When pJC28l was linearized at the HindIII site upstream of the tac promoter or at the NcoI, AatII, or PstI sites downstream of the 5B ORF (see Fig. 3), the 8.4-kDa product was still synthesized in large amounts (data not shown). Plasmids linearized at the SacII or BstEII sites within the 5' end of the 5B ORF, however, failed to yield detectable levels of this product. Furthermore, when pJC28l was digested at the NsiI or StyI sites further into the ORF, truncated polypeptides of increasing size were observed (approximately 4-5 kDa and 5-6 kDa, respectively). This data is consistent with the argument that the 8.4-kDa protein is expressed from the 5B ORF and inconsistent with the formal possibility that it is encoded either by the opposite strand of the DNA insert or by a fusion of 5B sequences with external vector sequences. Thus, this product will hereafter be referred to as the 5B protein.

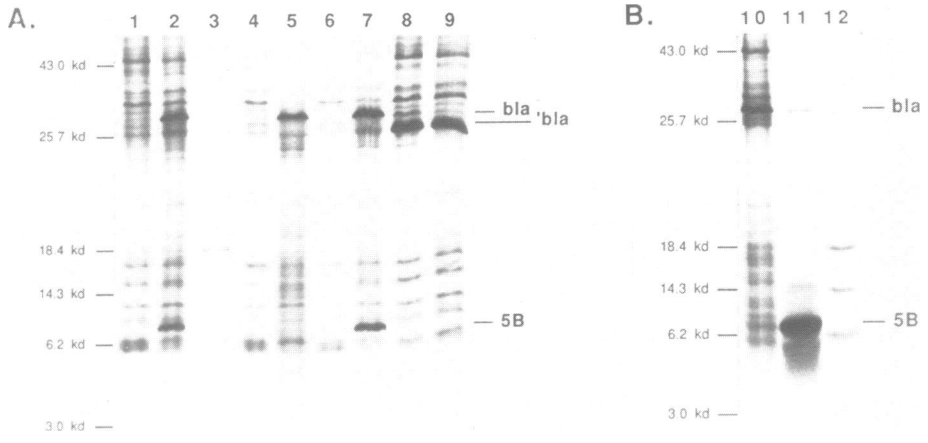


FIG. 5. In vivo synthesis and degradation of the 5B protein in cells containing pJC281. RB791 laci^q cells containing the plasmids designated below were either pulse-labelled with [^{3}H]leucine for 30 sec (lanes 1,2,4-7,10) or labelled continuously for 30-50 min (lanes 8-9). Where indicated, labelling reactions followed induction with IPTG for the time period shown. Lane 11 contains proteins synthesized in vitro using 50 µCi of [^{3}H]leucine. Samples were prepared for electrophoresis, analyzed on a 20% SDS-polyacrylamide gel cross-linked with DATD, and visualized by autoradiography. Lanes: 1, pJC281, 60 min without induction; 2, pJC281, 60 min induction; 3, molecular weight markers; 4, ptac 11, 120 min without induction; 5, ptac 11, 120 min induction; 6, pJC281, 120 min without induction; 7, pJC281, 120 min induction; 8, ptac 11, 45 min induction and 30-50 min labelling; 9, pJC281, 45 min induction and 30-50 min labelling; 10, pJC281, 30 min induction; 11, pJC281, 40 min in vitro labelling; 12, molecular weight markers. The positions of bands corresponding to ß-lactamase (bla), the processed form of ß-lactamase ('bla), and the 5B protein (5B) are noted on the right.

Expression and degradation of the 5B protein in vivo

To measure the rate of synthesis of the 5B protein in vivo, E. coli RB791 laci^q cells containing ptac11 or pJC281 were induced with IPTG and then pulse-labelled for 30 sec. As shown in Figure 5A, a protein of apparent molecular weight 8.4 kDa was produced from pJC281 at a rate of 20-25% of the total protein synthesis. Since ptac11 failed to make this protein, its production requires the presence of the 5B ORF. Furthermore, its appearance is dependent upon IPTG induction, indicating that it is under tac promoter control. When this 8.4-kDa species was electrophoresed on an SDS-polyacrylamide gel alongside the 5B protein

FIG. 6. Structure of the native 5B translation initiation region. A portion of the probable mRNA transcript for the 5B protein is shown in single-stranded stem-loop form (10). (A). The previously-suggested Shine-Dalgarno sequence (10) is given in boldface type and an alternative Shine-Dalgarno sequence is designated by the thin bar below and to the right of the appropriate nucleotides. Positions which can base pair with a critical region of the 3' end of 16S rRNA (5'-ACCUCC) are indicated by asterisks (to the left of the relevant nucleotides) in the case of the sequence suggested previously, or by crosses in the case of the alternative sequence suggested here. Recognition of either of these signals could lead to translation initiation at the 5B AUG start codon shown in boldface type. The alternative sequence might be preferentially utilized because it has greater potential for base pairing with the 16S rRNA outside of the critical region (not shown), and because the two nucleotides at its 5' end are not contained within the stem of the stem-loop structure. The circled C residue corresponds to a position which was changed to a G residue in the SRBS (see Fig. 3). (B). A third Shine-Dalgarno sequence and its associated UUG start codon are indicated in boldface type. Positions which can base pair with the 16S rRNA sequence mentioned above are designated by crosses to the right of the relevant nucleotides.

produced *in vitro*, the two were shown to comigrate (Fig. 5B).

Cells containing pJC281 which were induced with IPTG for 45 min and then labelled for 30-50 min did not accumulate detectable levels of 5B protein (Fig. 5A). This is in contrast to the results obtained with a 30-40 min labelling reaction *in vitro* (see Fig. 4). When induced cells containing pJC281 were pulse-labelled for 30 sec and then chased for 10 min (see Materials and Methods), again there was no detectable 5B protein (data not shown). This data strongly suggests that the 5B protein is degraded within 10 min *in vivo*.

DISCUSSION

We have described the *in vitro* and *in vivo* overproduction of the 5B protein of IS5 using an efficient synthetic ribosome binding site and a strong *tac* promoter. This protein was not detected

from a construct containing the native ribosome binding site and a tac promoter, suggesting that expression of the 5B gene is normally limited by poor translation initiation. ß-galactosidase fusion studies have indicated that the SRBS is at least 150-fold more efficient than the native ribosome binding site.

Examination of the native 5B translation initiation region suggests several reasons for its inefficiency. First, both the previously-suggested Shine-Dalgarno sequence (10) and a second one suggested here lack a significant stretch of polypurines and are only complementary to the 3' end of 16S rRNA at 3 out of 6 important positions (Fig. 6A). Second, both the Shine-Dalgarno sequences and the AUG start codon are likely to be unfavorably sequestered within the stem of a stable stem-loop structure (37,38). In addition, the spacer regions just upstream of the start codon contain 3-4 unfavorable G residues (32,33,35,38).

Another possible Shine-Dalgarno sequence, which is located just 2 bp upstream of the 5B AUG start (see Fig. 6B), may actually interfere with appropriate translation initiation. Unlike the others, it has a very good polypurine tract and is complementary to the 16S rRNA at 5 out of 6 positions, but it is inappropriately spaced from the AUG (33). Instead, it is favorably spaced from a possible UUG start codon which is in a different ORF. Since this UUG is not sequestered in any secondary structure and the strong Shine-Dalgarno sequence associated with it is only partially sequestered at its 3' end, this set of translation initiation signals may competitively inhibit binding of the ribosome to the appropriate 5B Shine-Dalgarno sequence. Initiation at this UUG, however, can only lead to production of a 3-kDa polypeptide which is unrelated to the 5B protein.

Although the 5B gene product is predicted to be 11.1 kDa in size, the observed protein has an apparent molecular weight of 8.4 kDa. The accidental introduction of a DNA mutation into the 5B ORF during the plasmid construction, which could lead to an early stop codon and thus a smaller protein, has been ruled out by sequencing the entire ORF of the final construct (12). Although the possibility of mRNA degradation cannot be completely excluded, in vitro experiments carried out with an RNase inhibitor or RNase$^-$ S30 extracts gave the same size polypeptide (data not shown,12). Similarly, since addition of protease inhibitors

or excess exogenous protein did not alter the size of the product, there is currently no evidence to suggest that this discrepancy is due to a nonspecific protease activity. Furthermore, *in vitro* reactions as short as 2 min (Fig. 4B) and *in vivo* labellings as short as 30 sec (Fig. 5) gave no evidence of a higher molecular weight 5B precursor. Labelling experiments with individual amino acids that are expected to occur only near the 5' end or the 3' end of the 5B protein indicated that at least residues 20 through 80 of the predicted 108-amino acid protein were present (data not shown). This data, as well as the sizes of the truncated products synthesized from *Nsi*I-digested and *Sty*I-digested pJC281 (see Results), is inconsistent with the formal possibility that the 8.4-kDa polypeptide is the result of translation initiation at a downstream ATG. One explanation for the discrepancy between the expected and observed sizes of the 5B protein is that of anomalous migration on SDS-polyacrylamide gels. An IS*4*-encoded protein with an apparent molecular weight significantly less than that predicted has also been reported (6), and a small regulatory protein encoded by the P2 bacteriophage *ogr* gene has a difference in gel migration similar to that seen here (42). In these cases as well, there was no clear explanation for the observed discrepancies.

At this point the nature of the discrepancy between the 12.3-kDa 5B gene product identified by Rak et al. (4) and the 8.4-kDa 5B protein seen here is unclear. Even though the IS*5* element utilized in the earlier study was isolated independently of that used here, the DNA sequences of the two are identical (10,11), and the electrophoretic conditions used for estimation of the molecular weight of the protein were similar. Although it is most likely that the 12.3-kDa product originates from the ATG start at position 525 (4,10), there is another ATG further upstream (see Fig. 1) which could have been utilized in the previous study but is not present in our pJC281 construct. Translation from this upstream start could lead to an expected product (in their case) of about 12.2 kDa rather than 11.1 kDa. However, use of the functional promoter just upstream of the 5B ORF (4) would result in a transcript which does not contain enough bases upstream of this start to allow for a ribosome binding site.

Thus the previous authors suggested that translation began at position 525, and we chose to overexpress the 5B ORF starting at this position. The apparent difference in size between the previously-observed product and that seen here may be related to the fact that the previous workers detected low level synthesis of a 5B protein in minicells while the current results were obtained with a 5B-overproducing plasmid *in vitro* and in whole cells. There have been several cases reported in which different levels of proteins or entirely different products were detected in minicells than *in vitro* or in whole cells (43,44), including that of an IS*4*-encoded product which was produced in low amounts in minicells but not at all *in vitro* (6). It remains to be shown whether the 8.4-kDa 5B gene product or the 12.3-kDa 5B gene product corresponds to a protein normally made *in vivo*.

The 5B protein is stable for at least 30-40 min *in vitro* but appears to be degraded within 10 min *in vivo*. Thus expression of the 5B gene may be maintained at a low level by protein degradation as well as poor translation initiation. Since the *in vitro* yield of 5B protein is significantly decreased when an S30 extract from an $RNase^+$ strain rather than an $RNase^-$ strain is used (12), RNA instability may also limit the level of 5B protein normally present *in vivo*.

Although we currently have no direct evidence to suggest a physiological role for the 8.4-kDa 5B protein, it seems highly likely that an ORF utilizing 354 bp of an 1195-bp insertion sequence would have an important role to play in either the regulation or mechanism of that element's transposition. This ORF has probably been maintained over a long period of time, and is similar in size and location to that of several other insertion sequences (2,8). In the case of Mu, Tn*5*, Tn*3* and gamma-delta, it is already known that two proteins are involved in maintaining wild type levels of transposition (1), and this appears to be true for IS*1* and IS*2* as well (45,46). By analogy to these other systems, one might imagine that the larger IS*5* 5A gene encodes a transposase and the smaller IS*5* 5B gene encodes either a regulatory protein or an auxiliary factor important to the transposition process itself.

This work has demonstrated that expression of the 5B gene is

strongly limited by poor translation initiation and that the 5B protein itself is unstable in whole cells. Previous studies have shown that the activity of the 5B promoter is very weak *in vivo* (4) and the 5B mRNA may be unstable (12). While the 5A promoter is somewhat stronger (4), the 5A protein is still produced in small amounts (3,4) and appears to be degraded into one or more lower molecular weight polypeptides (3). The gene products encoded by other bacterial transposable elements also appear to be maintained at low levels (1,5,6,47). In particular, protein instability has been demonstrated or suggested for the IS*903*, IS*10*, and bacteriophage Mu transposases (48-50), and it is interesting to note that the translation initiation regions of several insertion sequence genes suggest their involvement in determining levels of protein production (10,51,52). It is likely that a strategy of multiple controls of gene expression has been selected for during the evolution of many transposable elements, since their propagation is dependent upon maintaining a transposition rate which is low enough to allow survival of the host.

ACKNOWLEDGEMENTS

We thank Bernard Weiss and Randall Reed for discussion, Srinivasan Chandrasegaran for assistance with DNA sequencing, and John Lee for his contributions to the ß-galactosidase studies.

This work was supported by U.S. National Institutes of Health training grants 5-T32-CA09139 and 5-T32-GM07445 to J.M.C. and by Public Health Service research grant 5-PO1-CA16519. H.O.S. is an American Cancer Society Research Professor.

*To whom correspondence should be addressed

+Present address: E.I. du Pont de Nemours and Co., Inc., Central Research and Development Department, Experimental Station E328/150, Wilmington, DE 19880-0328, USA

REFERENCES

1. Craig, N.L. and Kleckner, N. (1987) In Neidhardt, F.L. (ed), *Escherichia coli* and *Salmonella typhimurium*, American Society for Microbiology, Washington, D.C., pp. 1054-1070.
2. Iida, S., Meyer, J. and Arber, W. (1983) In Shapiro, J.A. (ed), Mobile Genetic Elements, Academic Press, New York, pp. 159-221.
3. Chernak, J.M., Schlaffer, E.J. and Smith, H.O. (1988) J. Bacteriol. *170*, 5368-5370.

4. Rak, B., Lusky, M. and Hable, M. (1982) Nature 297, 124-128.
5. Rothstein, S.J., Jorgensen, R.A., Yin, J.C.-P. Yong-di, Z., Johnson, R.C. and Reznikoff, W.S. (1981) Cold Spring Harbor Symp. Quant. Biol. 45, 99-188.
6. Trinks, K., Habermann, P., Beyreuther, K., Starlinger, P. and Ehring, R. (1981) Mol. Gen. Genet. 182, 183-188.
7. Zerbib, D., Jakowec, M., Prentki, P., Galas, D.J. and Chandler, M. (1987) EMBO J. 6, 3163-3169.
8. Rak, B. and von Reutern, M. (1984) EMBO J. 3, 807-811.
9. Engler, J.A. and van Bree, M.P. (1981) Gene 14, 155-163.
10. Kröger, M. and Hobom, G. (1982) Nature 297, 159-162.
11. Schoner, B. and Kahn, M. (1981) Gene 14, 165-174.
12. Chernak, J.M. (1987) Ph.D. Thesis, The Johns Hopkins University, Baltimore, MD.
13. Boyer, H.W. and Roulland-Dussoix, D. (1969) J. Mol. Biol. 41, 459-472.
14. Amann, E., Brosius, J. and Ptashne, M. (1983) Gene 25, 167-178.
15. Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) Gene 33, 103-119.
16. Talmadge, K. and Gilbert, W. (1980) Gene 12, 235-241.
17. Bachmann, B.J. (1972) Bacteriol. Rev. 36, 525-557.
18. Silhavy, T.J., Berman, M.L. and Enquist, L.W. (1984) Experiments with Gene Fusions, Cold Spring Harbor, New York.
19. Miller, J.H. (1972) Experiments in Molecular Genetics, Cold Spring Harbor, NY.
20. Bolivar, F., Rodriguez, R.L., Greene, P.J., Betlach, M.C., Heyneker, H.L.,Boyer, H.W., Crosa, J.H. and Falkow, S. (1977) Gene 2, 95-113.
21. Kelly, S. and Smith, H.O., personal communication.
22. Russell, D.R. and Bennett, G.N. (1982) Gene 20, 231-243.
23. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor, NY.
24. Grunstein, M. and Hogness, D.S. (1975) Proc. Natl. Acad. Sci. USA 72, 3961-3965.
25. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
26. Sancar, A., Wharton, R.P., Seltzer, S., Kacinski, B.M., Clarke, N.D. and Rupp, W.D. (1981) J. Mol. Biol. 148, 45-62.
27. Laemmli, U.K. (1970) Nature 227, 680-681.
28. Bonner, W.M. and Laskey, R.A. (1974) Eur. J. Biochem. 46, 83-88.
29. Laskey, R.A. (1980) Meth. Enzymol. 65, 363-371.
30. DeVries, J.K. and Zubay, G. (1967) Proc. Natl. Acad. Sci. USA 57, 1010-1012.
31. Wilson, D.B. and Hogness, D.S. (1966) Meth. Enzymol. 8, 229-235.
32. Scherer, G.F., Walkinshaw, M.D., Arnott, S. and Morre, D.J. (1980) Nucl. Acids Res. 8, 3895-3907.
33. Stormo, G.D., Schneider, T.D. and Gold, L.M. (1982) Nucl. Acids Res. 10, 2971-2996.
34. Jay, G., Khoury, G., Seth, A.K. and Jay, E. (1981) Proc. Natl. Acad. Sci. USA 78, 5543-5548.
35. deBoer, H.A., Hui, A., Comstock, L.J., Wong, E. and Vasser, M. (1983) DNA 2, 231-235.

36. Hui, A., Hayflick, J., Dinkelspiel, K. and deBoer, H.A. (1984) EMBO J. 3, 623-629.
37. Buell, G., Schulz, M.F., Selzer, G., Chollet, A., Movva, N.R., Semon, D., Escanez, S. and Kawashima, E. (1985) Nucl. Acids Res. 13, 1923-1938.
38. Stormo, G.D. (1986) In Reznikoff, W. and Gold, L. (eds), Maximizing Gene Expression, Butterworth, Stoneham, MA, pp. 195-224.
39. Grantham, R., Gautier, C., Gouy, M., Jacobzone, M. and Mercier, R. (1981) Nucl. Acids Res. 9, r43-r74.
40. Grosjean, H. and Fiers, W. (1982) Gene 18, 199-209.
41. Ikemura, T. (1981) J. Mol. Biol. 151, 389-409.
42. Birkeland, N.K. and Lindqvist, B.H. (1986) J. Mol. Biol. 188, 487-490.
43. Collins, J. (1979) Gene 6, 29-42.
44. Johnson, R.C., Yin, J.C.P. and Reznikoff, W.S. (1982) Cell 30, 873-882.
45. Jakowec, M., Prentki, P., Chandler, M. and Galas, D.J. (1988) Genetics 120, 47-55.
46. Musso, R.E., personal communication.
47. Raleigh, E.A. and Kleckner, N. (1986) Proc. Natl. Acad. Sci. USA 83, 1787-1791.
48. Derbyshire, K.M. and Grindley, N.D.F., personal communication.
49. Morisato, D., Way, J.C., Kim, H-J. and Kleckner, N. (1983) Cell 32, 799-807.
50. Pato, M.L. and Reich, C. (1982) Cell 29, 219-225.
51. Halling, S.M., Simons, R.W., Way, J.C., Walsh, R.B. and Kleckner, N. (1982) Proc. Natl. Acad. Sci. USA 79, 2608-2612.
52. Klaer, R., Tillmann, E., Fritz, H.-J. and Starlinger, P. (1981) Mol. Gen. Genet. 181, 169-175.