

# A novel translation initiation region from *Mycoplasma genitalium* that functions in *Escherichia coli*

Steve Loechel<sup>1</sup>, Julia M. Inamine<sup>1,+</sup> and Ping-chuan Hu<sup>1,2,3,\*</sup>

Departments of <sup>1</sup>Pediatrics, <sup>2</sup>Microbiology and Immunology and <sup>3</sup>Center for Environmental Medicine, University of North Carolina, Chapel Hill, NC 27599-7220, USA

Received July 23, 1991; Revised and Accepted November 20, 1991

## ABSTRACT

The *tuf* gene of *Mycoplasma genitalium* uses a signal other than a Shine-Dalgarno sequence to promote translation initiation. We have inserted the translation initiation region of this gene in front of the *Escherichia coli lacZ* gene and shown that it is recognized by the translational machinery of *E. coli*; the signal operates *in vivo* at roughly the same efficiency as a synthetic Shine-Dalgarno sequence. The *M. genitalium* sequence was also used to replace the native translation initiation region of the *cat* gene. When assayed in *E. coli*, the *M. genitalium* sequence is equivalent to a Shine-Dalgarno sequence in stimulating translation of this mRNA also. Site-directed mutagenesis enabled us to identify some of the bases that comprise the functional sequence. We propose that the sequence UUAACAA-CAU functions as a ribosome binding site by annealing to nucleotides 1082–1093 of the *E. coli* 16S rRNA. The activity of this sequence is enhanced when it is present in the loop of a stem-and-loop structure. Additional sequences both upstream and downstream of the initiation codon are also involved, but their role has not been elucidated.

## INTRODUCTION

It is generally accepted that the efficient translation of bacterial mRNAs depends on the presence of a Shine-Dalgarno (SD) sequence 5' to the initiation codon. This sequence, 5'-AGG-AGG-3' or a subset thereof, anneals to the 3' end of the 16S ribosomal RNA. The involvement of the SD sequence in translation initiation has been well-studied, particularly in *E. coli* (1,2,3,4). There are a few exceptional mRNAs that contain no SD sequence or a very poor one, but these mRNAs are also poorly translated.

Not surprisingly, use of the SD sequence also appears to be widespread in various mycoplasma species, whose genes have begun to be sequenced in the past few years (5,6,7,8). We have encountered an interesting case in *Mycoplasma genitalium*, however, where a highly expressed gene has no SD sequence preceding the initiation codon. Elongation factor Tu, encoded

by the *tuf* gene, is a highly conserved protein used for translation; it is one of the most abundant proteins in the prokaryotic cell (9). We have previously reported the sequencing of the single *tuf* gene of *M. genitalium* (10). The gene is very similar to the homologous genes found in *E. coli* as well as other bacteria, except that there is no SD sequence, even a weak one, upstream of the AUG initiation codon. There is also no open reading frame immediately preceding the *tuf* gene, so that translational coupling to another gene with a SD sequence is not a possibility. Yet somehow the mRNA of this gene must be efficiently translated in order to supply the large amount of elongation factor Tu needed for protein synthesis.

In order to determine which sequences in the *tuf* mRNA are responsible for stimulating translation initiation, a number of techniques are required that are presently not available for use in *M. genitalium*. We therefore chose to study this translation initiation region (TIR) in *E. coli*, in the hope that the *E. coli* translational machinery would be able to recognize the signal sequence present in the *M. genitalium* mRNA. Preliminary experiments encouraged this approach, for they showed that *E. coli* recognized the sequence at the beginning of the *tuf* gene as a translation initiation site.

We have replaced the TIR of the *E. coli lacZ* gene with its counterpart from the *M. genitalium tuf* gene. In another set of experiments, different portions of the *M. genitalium* TIR were used to replace the TIR from the Tn9 gene coding for chloramphenicol acetyltransferase (CAT). Mutagenesis of these constructs and measurement of their translational efficiencies allowed us to determine which nucleotides of the *M. genitalium* sequence are required for stimulation of translation, and also to test our hypothesis for the mechanism by which this stimulation occurs.

## MATERIALS AND METHODS

### Bacteria, plasmids, and phage

*E. coli* strain DH5 $\alpha$ F'IQ from Bethesda Research Laboratories was used for propagation of M13 phage. *E. coli* strain JM109 (11) was used for growing plasmids. Plasmid pUR290 has been

\* To whom correspondence should be addressed

+ Present address: Department of Microbiology, Colorado State University, Fort Collins, CO 80523, USA

previously described (12). We obtained pUC-4K and pCM-1 from Pharmacia LKB, pIBL24 from International Biotechnologies, Inc., and pBR322, M13mp18, and M13mp19 from Bethesda Research Laboratories. Cells were grown at 37°C in YT medium (10g bactotryptone, 10g yeast extract, 5g NaCl per liter), and plating of M13 phage to obtain blue or white plaques was done as described (13).

### Oligonucleotides

Oligonucleotides were synthesized in the laboratory of Clyde Hutchinson III or purchased from National Biosciences. The following oligonucleotides, with their sequences written 5' to 3', were used: CAGGGTCAGAGCTGTTTGGTGTGTG for mutagenesis of M13mp19. AAAACGACTGCCAATGAATTC for mutagenesis of clone 120-3. GCCATTTGTTAAATTATGTTGTTAATTTCCAC, a degenerate oligonucleotide for saturation mutagenesis of clone 124-1. Positions 1–5 and 29–34 are pure, while positions 6–28 contain 91% of the listed base and 3% each of the unlisted bases. To make the specific mutations shown in Figure 4, an oligonucleotide of the last sequence shown above was synthesized, with the appropriate base changes. AG-ATCGCACTCCAGCCAGC, which anneals to the *lacZ* gene downstream of the *FspI* site, was used to confirm the sequence of the *AhaII/FspI* fragment before transferring it to a plasmid.

An M13 24mer universal sequencing primer was purchased from New England Biolabs for S1 nuclease mapping of the transcription start site and for quantitation of mRNA levels in the *lacZ* constructs. AAGCTGTAAACTAGAGAATTTCG for introducing stop codons in the open reading frames upstream of the *cat* gene. GCCTCAAATGTTCTTTACGATGCC for quantitation of *cat* mRNA. The following oligonucleotides were used for mutagenesis of the *cat* TIR:

GAGATTTTCATTAGCTAATTAAGCTAA  
GAGATTTTCTTAACAACATTAAGCTAA  
TTCTTAACCCAACATTAAGC  
CTTAACAACATAATTTAAACAAATGGAG  
GACGAGATTTGCTTAACCCAACAT  
TTTCGACGAGATTAATTAACAACATAATT  
GAGAATTTGACCACATTAATTAACAAC

### Nucleic acid techniques

Plasmids and phage clones were constructed by standard techniques (13). DNA sequencing was done by the dideoxy chain-termination method of Sanger (14).

Deletions of clones 118-3 and 119-13 were done by the Exonuclease III method of Henikoff (15) using the Erase-a-base system from Promega Biotech. Digestion was performed at 20°C in order to slow the reaction rate.

Site-directed mutagenesis using oligonucleotides was performed by the method of Kunkel (16), using a kit purchased from Bio-Rad. A two- to ten-fold molar excess of primer to template was used, except in the case of saturation mutagenesis, where equimolar amounts of primer and template were used. All clones with mutations of interest were plaque-purified and resequenced.

S1 nuclease mapping was done as described (13).

Analysis of nucleotide sequences was performed using the Beckman Microgenie program. Prediction of secondary structure was based on the rules formulated by Tinoco *et al.* (17).

A portion of the *M. genitalium* 16S rRNA gene was sequenced by using the polymerase chain reaction as described by Edwards *et al.* (18).

### Quantitation of $\beta$ -galactosidase activity

Translation initiation regions from each M13 clone were transferred to a plasmid containing the 3' portion of the *lacZ* gene. A 400 bp *AhaII/FspI* fragment from the M13 clones contains the *lac* promoter, the mutated translation initiation region, and the beginning of the *lacZ* gene. The plasmid p75A4 was constructed as shown in Figure 2. The important features of p75A4 are a unique *BstBI* site for ligation to the *AhaII* site upstream of the *lac* promoter, and the presence of the 3' portion of the *lacZ* gene from the *FspI* site to the end, with no other *FspI* sites in the plasmid.

The entire 400 bp DNA fragment that was transferred from M13 clone to plasmid was sequenced for each clone, to ensure that no fortuitous mutations were introduced during site-directed mutagenesis. p75A4 is a high-copy-number plasmid, and the pGal clones that had moderate to high translational efficiency produced so much  $\beta$ -galactosidase that cell growth was substantially slowed. In order to avoid artifacts that could result from such abnormally high mRNA levels, we cotransformed the cells with pBR322. When the cells were grown in the presence of tetracycline and kanamycin, both pBR322 and the pGal plasmids were stably maintained. The Rop protein and wild-type RNA I encoded by pBR322 (19) caused at least a ten-fold reduction in the copy number of the pGal plasmids.

*E. coli* cells carrying the plasmids were grown to mid-log phase in YT medium containing 35  $\mu$ g/ml kanamycin and 15  $\mu$ g/ml tetracycline; the doubling time was approximately 32 minutes. Cells were harvested and assayed according to the method of Miller (20). Assays were done in duplicate on at least two independently grown cultures.

### Quantitation of CAT synthesis

An 800 bp *SalI* fragment of pCM-1 containing the *cat* gene and its TIR (21) was cloned in the *EcoRI* site of pBR322 by filling in the ends and then ligating. In the orientation where the beginning of the *cat* gene is adjacent to the *tet* gene, this puts the *cat* gene under control of the 'antitet' promoter, with transcription initiation occurring 71 nt upstream of the translation initiation codon (22). A 300 bp *EcoRI/HindIII* fragment containing the TIR was put in M13mp18 for site-directed mutagenesis. After mutagenesis, the 300 bp fragment was completely sequenced and put back in the plasmid to reconstitute the complete *cat* gene.

Cells were grown to mid-log phase in YT medium containing 100  $\mu$ g/ml ampicillin. Doubling time was approximately 27 minutes. Cells were harvested, suspended in cold 50 mM TRIS (pH 7.8) and lysed by sonication. CAT levels were measured using an immunoassay from 5 Prime-3 Prime. Protein concentration was determined using the Bio-Rad Protein Assay, with bovine gamma globulin as the standard. Assays were done on at least two independently grown cultures.

### Quantitation of mRNA levels

*E. coli* cells were grown exactly as they were for  $\beta$ -galactosidase or CAT assays, and total RNA was isolated (13). Equal amounts of RNA were bound to GeneScreenPlus membrane from NEN Research Products, using a Bio-Dot SF apparatus from Bio-Rad. An oligonucleotide complementary to the *lacZ* or *cat* mRNA was end-labelled with <sup>32</sup>P (13) and hybridized to the RNA. Binding of RNA to the membrane and hybridization to the probe were done as described in the GeneScreenPlus protocol. Bands on the autoradiograph were quantitated by densitometry.

**RESULTS**

**Delineating the boundaries of the *M. genitalium* TIR**

We chose to use the *E. coli lacZ* gene as a reporter for the translation initiation activity of the *M. genitalium* sequence. The phage M13mp19 contains the *lac* promoter and the first portion of the *lacZ* gene, with a polylinker containing multiple restriction endonuclease sites inserted just downstream of the *lacZ* initiation codon (23). When plated in the presence of the inducer isopropylthiogalactoside (IPTG) and the indicating dye 5-bromo-4-chloro-3-indolyl- $\beta$ -D-galactopyranoside (X-Gal), the *lacZ'* gene product is synthesized; it complements the C-terminal fragment of  $\beta$ -galactosidase produced by certain *E. coli* strains, resulting in a dark blue rather than a white plaque.

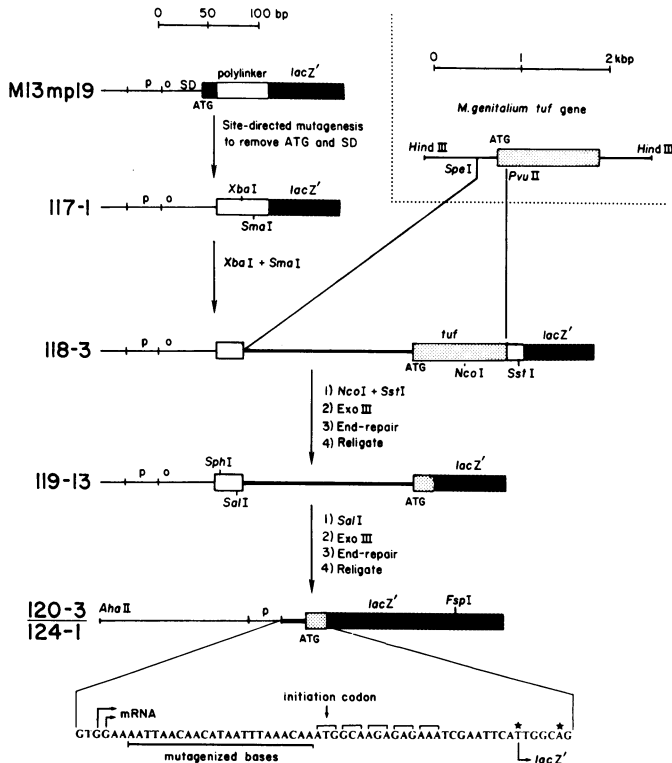
It was first necessary to destroy the *lacZ* translation initiation site in M13mp19, and we did so by site-directed mutagenesis. Both the SD sequence and the two possible ATG initiation codons were altered; as expected, the clone 117-1 (Fig. 1) gave white plaques.

We have reported elsewhere the cloning and sequencing of the *tuf* gene of *M. genitalium* (10). A 265 bp *SpeI/PvuII* DNA fragment containing the ATG initiation codon and 170 bp of noncoding sequence was inserted at the *XbaI/SmaI* sites of 117-1. Clone 118-3 contains the beginning of the *M. genitalium tuf* gene fused in-frame to the *lacZ'* gene, under control of the *lac* promoter and operator. It produces light blue plaques, indicating that *E. coli* recognizes the *tuf* initiation codon. (There are no other

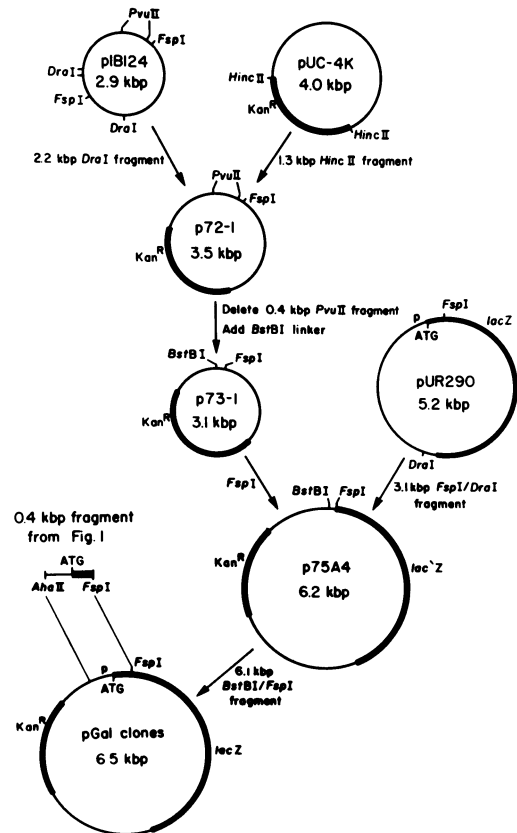
ATG or GTG codons in the portion of the *tuf* gene contained in 118-3.)

In order to narrow down the sequence responsible for activity, we used the enzyme exonuclease III (exo III) to selectively delete portions of the *M. genitalium* sequence. We first deleted excess *tuf* coding sequence as follows. The restriction endonuclease *SstI* cuts in the polylinker region remaining between the *tuf* and *lacZ'* genes. It leaves a 3' overhang, which is resistant to digestion by *exo III*. *NcoI* cuts within the *tuf* gene, and leaves a 5' overhang that is susceptible to *exo III* digestion. DNA from several digestion time points was end-repaired, religated, and transfected into *E. coli*. Clones that still gave blue plaques were isolated and sequenced. The clone with the largest deletion, 119-13, was selected. It retains only the first five codons of the *tuf* gene.

In order to delete *M. genitalium* DNA upstream of the initiation codon, 119-13 was cut at the *SalI* and *SphI* sites in the polylinker between the *lac* promoter and the *M. genitalium* insert. *Exo III* digested from the *SalI* site into the *M. genitalium* DNA, but the 3' overhang of the *SphI* site prevented it from deleting the promoter. DNA from several time points was religated and transfected, and clones that gave blue plaques were sequenced. The clones with the largest deletions retained 25 to 30 bases upstream of the initiation codon. One such clone, 120-3, also fortuitously deleted DNA in the opposite direction, possibly because it had failed to cut with *SphI*. 120-3 retains the *lac* promoter and the transcription start site of the wild-type *lacZ*



**Figure 1.** Construction of an M13 clone that contains the *lac* promoter, 41 bases of *M. genitalium* sequence with the initiation codon of the *tuf* gene, and the *E. coli lacZ'* gene. Transcription of clone 124-1 starts at the bases indicated and produces an mRNA whose translation is dependent on a signal sequence from *M. genitalium*. This clone was used as a template for making mutations in the *M. genitalium* sequence upstream of the initiation codon.



**Figure 2.** Construction of a plasmid vector for measuring the activity of mutant translation initiation regions. A DNA fragment from each mutated M13 clone was transferred to p75A4; this places the entire *lacZ* gene under control of the *lac* promoter and the *M. genitalium* translation initiation region.

gene (the fifth base in Fig. 1) (24), but the *lac* operator has been deleted. The initiation codon is preceded by 25 bases of *M. genitalium* sequence. We have determined by S1 nuclease mapping that the primary transcription start sites in this construct are the two G nucleotides immediately preceding the A which is used by the wild-type *lac* promoter (data not shown). Presumably the foreign DNA inserted downstream of the promoter is responsible for the slight shift in the initiation site.

The plaques produced by 120-3 are much darker blue than plaques produced by clones that retain the *lac* sequence deleted in 120-3. We do not know whether the increased activity is due to deletion of the operator, or to deletion of noncoding sequences from the mRNA that are deleterious to translation initiation.

The construct 120-3 therefore produces an mRNA with one or two G nucleotides at the 5' end, followed by 25 bases of noncoding *M. genitalium* sequence, an AUG initiation codon, 13 more bases of *tuf* coding sequence, and finally the *lacZ'* gene fused in-frame. The mRNA is predicted to be relatively unstructured at the 5' end. There is one strong predicted hairpin involving the initiation codon and the beginning of the *lacZ* gene. In order to eliminate this hairpin and the difficulties it could cause in interpreting the results of experiments, we introduced two silent mutations at the beginning of the *lacZ'* gene. The bases marked with a star in Figure 1 are the altered bases created by site-directed mutagenesis. This clone, 124-1, was used for further studies.

### Saturation mutagenesis of the TIR

Once we narrowed down the sequence responsible for stimulation of translation initiation, we could perform site-directed mutagenesis to determine exactly which nucleotides are involved. We chose to examine only the bases 5' of the initiation codon. An oligonucleotide was synthesized complementary to the sequence shown in Figure 1, from the first C downstream of the initiation codon to the beginning of the sequence. Those bases that anneal to the first 23 bases upstream of the initiation codon were doped with small amounts of mutant bases, so that a degenerate oligonucleotide was obtained that would create random mutations in this region. Site-directed mutagenesis was performed, and thousands of clones were plated on indicator medium. Clones were scored as giving dark blue, light blue, or white plaques; the wild-type clone 124-1 produces a dark blue

plaque. Less than one percent of the plaques were white or light blue. Clones from 83 white or light blue and 248 dark blue plaques were isolated and sequenced. The dark blue clones comprised 97 with no mutations, 84 with one, 45 with two, and 22 with three or more mutations. This is an expected result with the doping rate used for the mutagenic oligonucleotide.

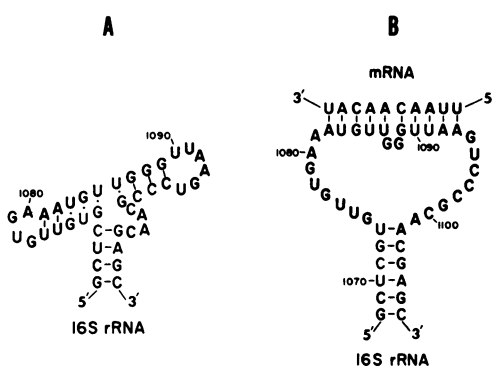
Many of the clones that produced white plaques were the result of fortuitous mutations, e.g. frame-shift mutations or mutations in the promoter. We obtained 39 clones that had mutations in the region of interest and produced white or light blue plaques.

The results of the saturation mutagenesis were inconclusive. Mutations at each of the 23 positions were found in clones that produce dark blue plaques, showing that no position is absolutely essential for activity. Conversely, the mutations in clones with reduced activity were scattered throughout the region, although mutations at some positions did seem more likely to result in reduced activity. All clones with reduced activity had two or more mutations. This experiment therefore did not enable us to narrow down the signal sequence any further.

### Testing a hypothetical mechanism

A reasonable mechanism for this *M. genitalium* TIR could involve annealing of a sequence to the 16S rRNA. We compared the sequence of the *M. genitalium tuf* translation initiation region to the *E. coli* 16S rRNA sequence. The most likely candidates for complementary sequences are nt 1082–1093 of the 16S rRNA and nt –12 to –21 of the *tuf* mRNA. They are exactly complementary, when a two-base bulge loop is allowed in the 16S rRNA sequence (Fig. 3). This region of rRNA sequence is highly conserved in all prokaryotes that have been examined, including many mycoplasma species (25,26,27).

We decided to test this hypothetical signal sequence by making specific mutations in the same construct that was used for saturation mutagenesis. Figure 4 shows the wild-type *M. genitalium* sequence, with the bases that are complementary to the 16S rRNA printed in large letters. Clone 117 from the saturation mutagenesis experiment has two mutations which reduce the complementarity, and this clone also has reduced  $\beta$ -galactosidase activity. If this TIR functions by annealing to the 1090 region of the 16S rRNA, then mutations to clone 117 which restore the complementarity of the 'anti-1090 sequence' to the 16S rRNA should increase the efficiency of translation initiation.



**Figure 3.** Sequence of the 1090 region of the *E. coli* 16S rRNA, shown in panel A with its predicted secondary structure (26). Panel B shows the hypothetical annealing of the 1090 region of the 16S rRNA and the anti-1090 region of the *M. genitalium tuf* mRNA signal sequence.

Clone	mRNA sequence	$\beta$ -gal. activity	mRNA level	translational efficiency
wild-type	ggaaaa <u>uuua</u> CAACUAAUUUUAAACAA <u>aug</u>	100	1.0	100
135-2 (SD)	ggaaaa <u>ucau</u> uauuAGGAGUAAACAA <u>aug</u>	144	1.0	144
117	ggaaaa <u>uca</u> CAACUAAUUUUAAACAA <u>aug</u>	22	2.4	9
136-1	ggaaaa <u>uaccca</u> CUAAUUUUAAACAA <u>aug</u>	170	2.7	63
140-10	ggAA <u>CUAA</u> ACCCAACAUAAUUUUAAACAA <u>aug</u>	138	1.1	125
16S rRNA	3'-UGAAUUGGGUUGUA-5' 1095 1082			

**Figure 4.** Mutations that were made in the untranslated leader sequence of the mRNA. The 5' terminus of the mRNA is at the left, and the initiation codon is underlined. For clone 135-2, the bases that comprise an SD sequence are in large letters. For other clones, bases that are complementary to the 1090 region of *E. coli* 16S rRNA are in large letters.  $\beta$ -galactosidase activity is normalized to that of the wild-type clone, which was 116 Miller units. Translational efficiency was calculated by dividing  $\beta$ -galactosidase activity by the level of mRNA, relative to wild-type.

To provide a standard for comparison, we created another construct which replaces the hypothetical signal sequence by an SD sequence.

Site-directed mutagenesis was performed to give the desired sequences, and  $\beta$ -galactosidase activity was measured as described in Materials and Methods. We also measured the level of *lacZ* mRNA in the cells, relative to the wild-type clone. This allows a better measure of the efficiency of translation initiation, since it allows us to correct for differences in transcription rates or mRNA stability.

Clone 135-2, with a synthetic SD sequence, has a translational efficiency 44% greater than the wild-type *M. genitalium* sequence. Clone 117, from the saturation mutagenesis experiment, has a translational efficiency equal to 9% of wild-type. Two C nucleotides were inserted in clone 117 at the position where they could bond to the two G nucleotides that would otherwise be found in a bulge loop on the 16S rRNA strand. As Figure 4 shows, this mutation was able to reverse the effect of the two mutations in clone 117 that decreased complementarity; translational efficiency increased seven-fold, from 9% to 63% of wild-type. Further increasing the complementarity from 8 bases (clone 136-1) to 14 bases (clone 140-10) gave a more modest increase in efficiency, to a level approaching the efficiency of an SD sequence. A clone with a frame-shift mutation immediately downstream of the initiation codon (deletion of the A at position +6) gave only 1% of wild-type activity (data not shown); this provides confirmation that translation initiation is occurring at the AUG codon.

### Detailed analysis of the TIR

To further study this TIR, we decided to insert different portions of it in the TIR of a different reporter gene, the *cat* gene. As described in Materials and Methods, a plasmid was constructed which has the *cat* gene under control of a plasmid promoter 71 bp upstream. This produces an mRNA with an untranslated leader consisting of 42 bases of plasmid sequence, preceding the *cat* TIR containing two SD sequences. The construct contains open reading frames extending into the TIR from upstream. To eliminate the possibility of ribosomes initiating translation upstream and continuing into the *cat* gene, site-directed mutagenesis was performed to put stop codons in all reading

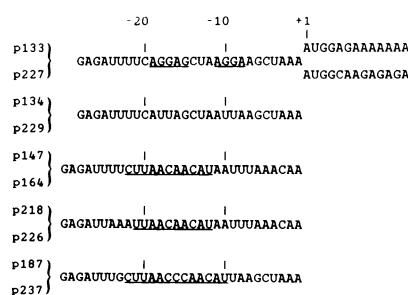
frames immediately upstream of the TIR. The sequence of the *cat* TIR is shown in Figure 5, as well as the sequences of the various mutants that were made.

CAT synthesis as well as mRNA levels were measured to determine the relative translational efficiency of each TIR (Table 1). As expected, destruction of the SD sequences decreased the efficiency drastically, to 3%. When nt -1 to -21 of the *M. genitalium* TIR were inserted upstream of the *cat* initiation codon (clone p147), we were surprised to see that there was no enhancement at all of translation initiation. This means that the anti-1090 sequence, located at -12 to -21, is not sufficient for activity; some part of the signal must be contained in the remainder of the TIR that was used in the *lacZ* constructs, namely nt -22 to -25 or +4 to +16.

Upon close examination of the predicted secondary structure of the TIR in the *lacZ* construct, we noticed that there is a weak hairpin with a free energy of -1.2 kcal/mol, with nt -7 to -11 annealing to nt -19 to -24 to form the stem, and the anti-1090 sequence being found in the loop. The *cat* construct p147 does not have nt -22 to -24 of the *M. genitalium* sequence, so it does not form this hairpin.

We altered nt -22 to -24 of p147 to create a hairpin with a free energy of -4.2 kcal/mol, with the anti-1090 sequence present in the loop (clone p218). The translational efficiency increased significantly, but still fell far short of the efficiency of the native TIR containing an SD sequence.

We next examined the effect of sequences downstream of the initiation codon. Five bases were changed so that the downstream sequence of these constructs matches the *M. genitalium tuf* sequence from +1 to +13. This had little effect on the efficiency of SD-mediated translation initiation, although it did increase the efficiency of the null construct from 3% to 7%. Changing the downstream sequence had a dramatic effect on the translational efficiency of clones that have the *M. genitalium* TIR sequence upstream of the initiation codon. The efficiency of the upstream sequence with no hairpin jumped from 3% to 33% (clones p147 and p164), and the efficiency of the upstream sequence with a hairpin increased from 20% to 100% of that of the native SD TIR (clones p218 and p226). It should be pointed out that the CAT assay we use is not based on enzymatic activity, but is an immunoassay employing a polyclonal antibody; changing the first



**Figure 5.** The translation initiation region of the *cat* gene, and mutants that were produced from it. The first base of the initiation codon is labelled +1. A pair of clones is listed for each TIR: the top one with *cat* sequence downstream of the initiation codon, the bottom one with *M. genitalium tuf* sequence downstream. Underlined bases for clones p133 and p227 mark the SD sequences of the *cat* TIR. Underlined bases for other clones mark the sequence that is complementary to the 16S rRNA 1090 region. The sequence of p238 is identical to that of p226, except that the first three bases are CAC rather than GAG.

**Table 1.** Efficiencies of various translation initiation regions

Clone	Upstream TIR	Downstream TIR	CAT level	mRNA level	Translational efficiency
p133	SD	<i>cat</i>	100.0	1.0	100
p134	null	<i>cat</i>	0.7	0.26	3
p147	<i>M. gen.</i> -1 to -21	<i>cat</i>	0.6	0.18	3
p218	<i>M. gen.</i> -1 to -21/hairpin	<i>cat</i>	6.1	0.31	20
p187	anti-1090/hairpin	<i>cat</i>	1.4	0.24	6
p227	SD	<i>M.gen.</i>	119.0	0.96	120
p229	null	<i>M.gen.</i>	1.2	0.17	7
p164	<i>M. gen.</i> -1 to -21	<i>M.gen.</i>	8.8	0.27	33
p226	<i>M. gen.</i> -1 to -21/hairpin	<i>M.gen.</i>	44.0	0.44	100
p237	anti-1090/hairpin	<i>M.gen.</i>	5.3	0.18	29
p238	p226, delete SD	<i>M.gen.</i>	55.0	0.51	110

Clone p133 has the native *cat* TIR. Destruction of the SD sequences produced the null construct p134, and then mutations were made to introduce different portions of the *M. genitalium* TIR. Sequences of these constructs are shown in Figure 5. CAT levels are normalized to that of p133, which produced 8.5  $\mu$ g CAT per mg total soluble protein. Translational efficiency was calculated by dividing CAT levels by the level of mRNA, relative to p133.

few codons should therefore have little effect on the measurement of CAT levels.

In order to determine whether the signal upstream of the initiation codon comprises only an anti-1090 sequence held in a hairpin, clones p187 and p237 were constructed. These contain an anti-1090 sequence, with two C nucleotides inserted to increase complementarity to the 16S rRNA. They contain a hairpin with a free energy of  $-5.4$  kcal/mol, with either *cat* sequence or *M. genitalium tuf* sequence at nt +1 to +13. These constructs do enhance translational efficiency, but to a far lesser extent than constructs with the entire *M. genitalium* upstream TIR.

There is a potential problem with interpreting the enhancement of translational efficiency seen with addition of a hairpin to the *M. genitalium* TIR. It is possible that the hairpin is functioning to bring an upstream SD sequence close enough to the initiation codon that it can function. In fact, clone p226, with translational efficiency equivalent to that of the native TIR, does bring a weak SD sequence (GAG) to within seven bases of the initiation codon. When this potential SD sequence was changed to CAC, there was no decrease in the activity of the TIR (clone p238). The hairpin therefore does not stimulate translation initiation by bringing the upstream SD sequence within range of the initiation codon.

In order to confirm that translation initiation in these clones is occurring at the first codon of the *cat* gene and not somewhere farther downstream, the nucleotide at position +14 was deleted to create a frame-shift. This abolished synthesis of the CAT protein (data not shown).

## DISCUSSION

We have shown that the translation initiation region of the *M. genitalium tuf* gene functions in *E. coli*, in spite of the fact that it contains no SD sequence. A plasmid was constructed that encodes an mRNA containing 25 bases of *M. genitalium* noncoding leader sequence, preceding an AUG initiation codon fused to the *lacZ* gene. This allowed us to measure the strength of the translation initiation signal by measuring  $\beta$ -galactosidase activity. An advantage of this construct is that it encodes an mRNA with a short untranslated leader sequence; since every additional base in the mRNA leader increases the complexity of the secondary structure of the TIR, a longer leader would make it that much more difficult to determine the role of primary sequences.

When saturation mutagenesis was performed on the TIR, the results implied that the mechanism by which it operates is complex. If a single specific sequence in the TIR were solely responsible for its activity, then the random mutations produced by saturation mutagenesis should have allowed its identification; no clearcut results were obtained, however. We were forced to take another approach to identify the mechanism by which the *M. genitalium* sequence stimulates translation initiation.

The translation initiation signal could conceivably function by interacting with a protein that recognizes its sequence. Alternatively, the 30S ribosomal subunit could recognize the sequence by base pairing of the 16S rRNA with the mRNA. The SD sequence functions in this way, and there is also evidence that the ribosome uses base-pairing between the 16S rRNA and mRNA to recognize UGA termination codons (28) and to recognize specific sequences that enhance the activity of an SD sequence (29,30).

The *M. genitalium* TIR contains the sequence UUAACAA-CAU, which is complementary to nt 1082–1093 of the *E. coli* 16S rRNA; we refer to it as the anti-1090 sequence. Base-pairing of these sequences could increase the efficiency of translation initiation by acting as a ribosome binding site. One point in favor of this mechanism is that the 16S rRNA sequence in question is highly conserved. This would explain how a signal from *M. genitalium* could function in such a distantly related organism as *E. coli*. We have sequenced this region of the *M. genitalium* gene coding for 16S rRNA; the sequence corresponding to nt 1082–1119 of the *E. coli* gene is exactly the same. In addition, it has been pointed out that there are two alternate secondary structures for this region of the 16S rRNA (26); one of these makes the proper nucleotides available for bonding to the mRNA. The 16S rRNA could exist in conformation B for binding to the mRNA signal sequence, and then shift to conformation A as translation commences (Fig. 3).

In order to test the above hypothesis, we made mutations in the functional sequence that increased complementarity to the 1090 region of the 16S rRNA. The insertion of two C nucleotides in the middle of the anti-1090 sequence compensated for the loss of complementarity caused by mutations at the ends of the sequence (Fig. 4). The wild-type anti-1090 sequence has 69% the efficiency of a synthetic SD sequence; the efficiency could be increased to 87% by increasing complementarity (clone 140-10).

Using the *lacZ* constructs, we showed that the *M. genitalium tuf* TIR functions nearly as well as a synthetic SD sequence. Using the *cat* constructs, we went on to show that the *M. genitalium* TIR is as efficient as a native TIR that uses an SD sequence as a ribosome binding site. In addition, the data from this set of experiments suggested that the *M. genitalium* TIR is made up of several components, not merely an anti-1090 sequence upstream of the initiation codon. There appear to be four components: 1) an anti-1090 sequence; 2) complementary sequences flanking the anti-1090 sequence to hold it in the loop of a hairpin; 3) a sequence at positions  $-1$  to  $-11$  of unknown function; 4) a sequence at positions  $+4$  to  $+13$  of unknown function. An anti-1090 sequence along with any two of the remaining components gives significant activity, ranging from 20% to 33% of the activity of a native TIR that uses an SD sequence (Table 1, clones p218, p164, p237). All four components together give 100% activity (clone p226).

It appears that increasing the length of the anti-1090 sequence can compensate for the absence of a hairpin, as long as the third and fourth components are present. In the *lacZ* constructs, clone 140-10 has a slightly higher efficiency than the wild-type *M. genitalium* sequence (Fig. 4); 140-10 has a 14-base anti-1090 sequence with no stable hairpin, while the wild-type clone has a 10-base anti-1090 sequence with a hairpin of free energy  $-1.2$  kcal/mol. It should be noted that the complete *M. genitalium* TIR has a slightly stronger hairpin of  $-2.4$  kcal/mol, due to an additional three bases of upstream sequence that were deleted in the *lacZ* constructs; the efficiency of the wild-type construct may have been lowered as a result. We have confirmed this aspect of the *M. genitalium* TIR by using the *cat* constructs. Clone p164 has three of the four components, but lacks a hairpin. When the length of the anti-1090 sequence in this construct is increased, CAT synthesis increases to the level seen with p226, which has all four components (data not shown).

We do not know what role nt  $-1$  to  $-11$  and nt  $+4$  to  $+13$  play in the mechanism of this TIR. Based on the comparison of

sequences as well as limited mutagenesis experiments, we feel it is unlikely that these bases are involved in mRNA-rRNA interactions or in the formation of hairpins or pseudoknots. It may be that these portions of the *M. genitalium* TIR are recognized and bound by ribosomal proteins rather than by the 16S rRNA.

Translation initiation of almost all known *E. coli* mRNAs occurs via a mechanism involving an SD sequence which anneals to the 3' end of the 16S rRNA. The efficiency of a TIR is dependent on the presence of additional sequences called translational enhancers. Olins and Rangwala have identified one such enhancer; according to their hypothesis, it anneals to nt 458–466 of the 16S rRNA (29). Sprengart *et al.* have identified another which is believed to anneal to nt 1471–1482 (30). These translational enhancers do not promote translation by themselves; they require the presence of an SD sequence in the TIR. This is in contrast to the *M. genitalium* sequence described in this paper, which functions in the absence of an SD sequence. There is no significant homology between the sequences of the *M. genitalium* TIR and the enhancers referred to above. Gallie and Kado have described a sequence from a eukaryotic mRNA which can function in place of an SD sequence in *E. coli*, although the efficiency was markedly lower (31). It is unclear whether there is any relationship between the sequence reported in their paper and the *M. genitalium* TIR.

Use of the SD sequence appears to be widespread among mycoplasma species, just as it is in all other bacteria that have been examined. The *tuf* genes of *M. gallisepticum* (32) and *M. capricolum* (unpublished data) have a clear-cut SD sequence preceding the initiation codon. In *M. genitalium* itself, the *mgp* operon has an excellent SD sequence preceding the first gene (6). As we have shown in this paper, the *M. genitalium tuf* TIR and a TIR based on an SD sequence can function at equivalent efficiencies *in vivo*; it would be of interest to learn whether one has any advantages over the other.

## ACKNOWLEDGEMENT

This work was supported by National Institutes of Health grant HL-19171 and by Environmental Protection Agency Cooperative Agreement CR807392.

## REFERENCES

- Gold, L. (1988) *Annu. Rev. Biochem.*, **57**, 199–233.
- Gren, E.J. (1984) *Biochimie*, **66**, 1–29.
- Shine, J., and Dalgarno, L. (1974) *Proc. Natl. Acad. Sci. USA*, **71**, 1342–1346.
- Stormo, G.D. (1986) In Reznikoff, W. and Gold, L. (eds.), *Maximizing Gene Expression*. Butterworths, Boston, pp. 195–226.
- Inamine, J.M., Loechel, S., and Hu, P.-c. (1988) *Gene*, **73**, 175–183.
- Inamine, J.M., Loechel, S., Collier, A.M., Barile, M.F., and Hu, P.-c. (1989) *Gene*, **82**, 259–267.
- Muto, A., Kawachi, Y., Yamao, F., and Osawa, S. (1984) *Nucleic Acids Res.*, **12**, 8209–8217.
- Renaudin, J., Pascarel, M.-C., and Bove, J.-M. (1987) *J. Bacteriol.*, **169**, 4950–4961.
- Pedersen, S., Bloch, P.L., Reeh, S., and Neidhardt, F.C. (1978) *Cell*, **14**, 179–190.
- Loechel, S., Inamine, J.M., and Hu, P.-c. (1989) *Nucleic Acids Res.*, **17**, 10127.
- Yanisch-Perron, C., Vieira, J., and Messing, J. (1985) *Gene*, **33**, 103–119.
- Rüther, U., and Müller-Hill, B. (1983) *EMBO J.*, **2**, 1791–1794.
- Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Smith, J.A., Seidman, J.G., and Struhl, K. (1987) *Current Protocols in Molecular Biology*. Wiley, New York.
- Sanger, F., Nicklen, S., and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
- Henikoff, S. (1984) *Gene*, **28**, 351–359.
- Kunkel, T.A. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 488–492.
- Tinoco, I., Jr., Borer, P.N., Dengler, B., Levine, M.D., Uhlenbeck, O.C., Crothers, D.M., and Gralla, J. (1973) *Nature New Biol.*, **246**, 40–41.
- Edwards, U., Rogall, T., Blöcker, H., Emde, M., and Böttger, E.C. (1989) *Nucleic Acids Res.*, **17**, 7843–7853.
- Minton, N.P., Chambers, S.P., Prior, S.E., Cole, S.T., and Garnier, T. (1988) *Focus*, **10**, 56.
- Miller, J.H. (1972) *Experiments in Molecular Genetics*. Cold Spring Harbor Laboratory, Cold Spring Harbor.
- Close, T.J., and Rodriguez, R.L. (1982) *Gene*, **20**, 305–316.
- Brosius, J., Cate, R.L., and Perlmutter, A.P. (1982) *J. Biol. Chem.*, **257**, 9205–9210.
- Norlander, J., Kempe, T., and Messing, J. (1983) *Gene*, **26**, 101–106.
- Reznikoff, W.S. (1984) *Symp. Soc. Gen. Microbiol.*, **36**, 195–218.
- Gutell, R.R., Weiser, B., Woese, C., and Noller, H.F. (1985) *Prog. Nucleic Acid Res. Mol. Biol.*, **32**, 155–216.
- Woese, C.R., Gutell, R., Gupta, R., and Noller, H.F. (1983) *Microbiol. Rev.*, **47**, 621–669.
- Weisburg, W.G., Tully, J.G., Rose, D.L., Petzel, J.P., Oyaizu, H., Yang, D., Mandelco, L., Sechrest, J., Lawrence, T.G., Van Etten, J., Maniloff, J., and Woese, C.R. (1989) *J. Bacteriol.*, **171**, 6455–6467.
- Hänfler, A., Kleuvers, B., and Göringer, H.U. (1990) *Nucleic Acids Res.*, **18**, 5625–5632.
- Olins, P.O., and Rangwala, S.H. (1989) *J. Biol. Chem.*, **264**, 16973–16976.
- Sprengart, M.L., Fatscher, H.P., and Fuchs, E. (1990) *Nucleic Acids Res.*, **18**, 1719–1723.
- Gallie, D.R., and Kado, C.I. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 129–132.
- Inamine, J.M., Loechel, S., and Hu, P.-c. (1989) *Nucleic Acids Res.*, **17**, 10126.