Sixteen single point mutations near the beginning of the lacZ gene have been isolated and their effect on lacZ expression has been measured. Five mutations were obtained that alter a potential stem-and-loop structure in the messenger RNA that masks the initiation codons. Formation of this stem-and-loop is a result of transcription of DNA sequences introduced during the cloning of the lac regulatory region. The mutations isolated were then moved into a background that deleted this structure. Analysis of these mutations indicated that the secondary structure inhibited lacZ expression 5.8-fold and that either single point mutations or a 9 base-pair deletion could relieve this inhibition completely. In addition, it was found that an A to C transversion in the first base following the initiation codon (in the absence of the inhibitory secondary structure) decreases lacZ expression almost twofold, whereas C to U transitions in the next two positions have negligible effects.

Mutations were also obtained that either increase or decrease the length of the Shine–Dalgarno sequence. The effects of these mutations were studied in the presence or absence of the secondary structure that involves the two initiation codons. It was found that when translation initiation was inhibited by the secondary structure, increasing the length of the Shine–Dalgarno sequence increased lacZ expression 2.8-fold and decreasing the length of this sequence reduced lacZ expression 12-fold. When translation initiation was not inhibited by the secondary structure, increasing the length of the Shine–Dalgarno sequence had no effect and decreasing the length of this sequence only reduced lacZ expression sixfold. The mechanistic implications of these results are discussed.

Two initiation codons are located in the beginning of the lacZ gene, 7 and 13 bases from the Shine–Dalgarno sequence. NH₂-terminal sequence analysis

† Present address: Chemistry Division, California Institute of Technology, Pasadena, Calif. 91125, U.S.A.
indicated that the majority of the protein synthesized initiate at the first initiation codon in the wild-type lacZ gene (in agreement with results reported previously by J. L. Brown and his colleagues). Upon introduction of sequences that result in a change in the mRNA secondary structure, both initiation codons are used in almost equal amounts. Three mutations and two pseudorevertants were obtained, which are located in the first initiation codon. It was found that when the first initiation codon is changed from AUG to GUG, translation initiation is decreased tenfold at that codon. Other mutations at the first initiation codon resulted in initiation occurring exclusively at the second initiation codon at an efficiency that is determined by the nucleotides located in the spacer region.

1. Introduction

Initiation of translation in procaryotes involves binding of the ribosome and the initiator transfer RNA to the messenger RNA, so that the message is appropriately positioned for translation (for a review see Kozak, 1983). The ribosome binding sites on the message were initially identified as regions on the mRNA protected by the ribosome from nuclease digestion (Steitz, 1969). The protected RNA generally extends 20 bases on either side of the initiation codon. Within this region, approximately five to nine bases preceding the initiation codon, a sequence is usually found that is complementary to a pyrimidine tract in the 3' end of 16 S ribosomal RNA. Shine & Dalgarno (1974) postulated that this sequence in the mRNA (frequently referred to as the Shine–Dalgarno sequence) could base-pair with the 16 S rRNA and thus serve as a recognition site for the ribosome.

The binding of the fMet-tRNA to the ribosome was originally believed to be mediated by codon-anticodon base-pairing with the previously bound mRNA (Haselkorn & Rothman-Denes, 1973; Benne et al., 1973). Though fMet-tRNA binding is stabilized by mRNA, more recent studies have shown that an initiation codon does not have to be present for this to occur (Jay et al., 1980; van der Laken et al., 1979; Dunn et al., 1978). It has been proposed that the binding of the mRNA to the 30 S subunit leads to a conformational change in the ribosome that stabilizes the binding of the fMet-tRNA (van der Laken et al., 1979; Gold et al., 1981). The interaction between the anticodon loop of the initiator tRNA and the initiation codon could then occur subsequently (Gualerzi et al., 1977).

The mRNA sequence could affect the formation of this ternary initiation complex in several ways. One could imagine that the length of the Shine–Dalgarno sequence may in part determine the stability of the mRNA in a 30 S–mRNA complex (or 30 S–tRNA–mRNA complex). In addition, both the sequence of the initiation codon (AUG, GUG or UUG) and the distance from the Shine–Dalgarno sequence to the initiation codon may affect the efficiency at which the initiator tRNA can recognize the initiation codon (Jay et al., 1981; Roberts et al., 1979). The ribosome may also recognize other signals in the mRNA or certain residues may decrease the stability of the message in the mRNA binding site. For instance, a computer analysis of ribosome binding sites revealed that three bases before the initiation codon an A residue is frequently found and the ribosome binding sites are generally deficient in G residues (Stormo et al., 1982). Finally,
any secondary structure in the mRNA that involves either the Shine–Dalgarno sequence or the initiator codon could prevent these interactions and thus lower the frequency of initiation (Saito & Richardson, 1981; Hall et al., 1982).

We report here the isolation and characterization of a series of mutations located in the lacZ ribosome binding site. The properties of these mutations provide insight into the effects on translation initiation of alterations in each of the sequence determinants mentioned above.

2. Materials and Methods

(a) Enzymes and reagents

Large fragment DNA polymerase I, restriction endonucleases EcoRI, SalI, AvaII and HindIII, and the indicator dye, 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (XG) were purchased from Bethesda Research Laboratories. S1 nuclease was obtained from Boehringer-Mannheim and bacteriophage T4 DNA ligase was a gift from R. Simoni. Universal primer was obtained from New England Biolabs. Isopropyl-β-D-thiogalactopyranoside was purchased from Bachem.

(b) Bacterial strains

Table 1 lists the bacterial strains used in this study along with the genotypes and references. To construct RZ203, RZ201 was transduced to Kan' with Plvir grown on CBK252. This strain (RZ202) was then transduced to iv' Δcyu with Plvir grown on CA8306. The Δcyu allele was checked by examining the phenotype of the resulting strain on maltose MacConkey plates. The Pl transducing lysates were a gift from A. Grossman.

(c) Construction of mRZ361 and mRZ352

mRZ361 was derived from M13mp8 and pRZ4024. pRZ4024 contains the HaeIII 203 bp lacPL8UV5 fragment between the EcoRI and SalI restriction sites of pBR322. An 840 bp AvaII-SalI restriction fragment was isolated and ligated to M13mp8 also digested with AvaII and SalI. The resulting vector was called mRZ36. The L8 mutation was separated from the UV5 mutations by digesting mRZ36 with EcoRI and SalI and ligating this vector to a EcoRI-HpaII lacP+ (CAP+) restriction fragment and a HpaII-SalI lacPUV5 restriction fragment. Introduction of the wild-type CAP site increased lacZ activity, allowing screening of the transfectants on indicator plates (top agar containing 0.3 mg XG/ml and 10⁻³ m-isopropyl-β-D-thiogalactopyranoside: Maniatis et al., 1982). The sequence was confirmed by DNA sequence analysis and this vector was named mRZ361 (see Fig. 1).

mRZ352 was obtained by digesting mRZ361 with SalI followed by S1 nuclease treatment for 40 min at 30°C (1.5 units S1 per μg of DNA in 100 mM-NaCl, 50 mM-sodium acetate (pH 4-0) and 6 mM-ZnSO₄). The DNA was then circularized with T4 DNA ligase and transfected intoJM101. Phage conferring an increase in lacZ activity were sequenced and 3 independent isolates contained the same 9 bp deletion (see Fig. 4).

(d) Mutagenesis and sequencing

mRZ361 was grown for 9 h in LB media on the strain RZ601. The replicative form of the phage was isolated and the EcoRI-SalI lacPUV5 restriction fragment was purified by polyacrylamide gel electrophoresis. This fragment was cloned into an unmutagenized

† Abbreviation used: bp, base-pairs.
<table>
<thead>
<tr>
<th>Strain</th>
<th>Genotype</th>
<th>Origin/reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>JM103</td>
<td>F' traD36 lacI*Zm15 proAB</td>
<td>Δ(lac-pro) supE thi strA endA</td>
</tr>
<tr>
<td></td>
<td>sbcB15 hsdR4 PI lysogen</td>
<td></td>
</tr>
<tr>
<td>JM101</td>
<td>F' traD36 lacI*Zm15 proAB</td>
<td>Δ(lac-pro) supE thi</td>
</tr>
<tr>
<td>RZ201</td>
<td>F' ara Δ(lac-pro) thi rpsL</td>
<td>Johnson et al. (1982)</td>
</tr>
<tr>
<td>RZ203</td>
<td>F' ara Δ(lac-pro) thi rpsL Δcya</td>
<td>This work</td>
</tr>
<tr>
<td>RZ411</td>
<td>pOX38-gen (F') ara Δ(lac-pro) nalA metB argEam rif supE</td>
<td>pOX38-gen (Johnson et al., 1982) transferred into XA102</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Miller et al., 1977)</td>
</tr>
<tr>
<td>RZ601</td>
<td>JM103 zaf13 :: Tn10 mutD5</td>
<td>P1vir transduction of KH1213 zaf13 :: Tn10 mutD5 (E. Cox) into JM103</td>
</tr>
<tr>
<td>XA103</td>
<td>F' ara Δ(lac-pro) nalA metB argEam rif supF</td>
<td>Miller et al. (1977)</td>
</tr>
<tr>
<td>CBK252</td>
<td>W3110 ilv :: Tn5</td>
<td>C. Berg</td>
</tr>
<tr>
<td>CA8396</td>
<td>Hfr Hayes thi rpsL Δcya</td>
<td>J. Beckwith</td>
</tr>
</tbody>
</table>
vector and phage conferring an increase or decrease in β-galactosidase activity were sequenced. The mutations isolated by this procedure were given numbers 1 to 11.

Alternatively, mRZ361 was treated in vitro with hydroxylamine for 10 min at 70°C (Humphreys et al., 1976). The mutagenized DNA was digested with EcoRI and SalI, and cloned into an unmutagenized vector. Mutations isolated by this method were numbered 30 to 37.

Mutations 76 and 77 were isolated as spontaneous mutations.

The DNA between the SalI site and the EcoRI site was sequenced by the dideoxy chain termination methods developed by Sanger et al. (1977). All of the mutant phage sequenced contained only one mutation in this region.

(e) Transfer of mutations into mRZ352

The transfer of the lacZ mutations into mRZ352 involved the hybridization of a 190 bp HindIII restriction fragment containing the lacPUV5 promoter and the mutation of interest to the single stranded phage, mRZ352-T743. (T743 is a strong promoter mutation introduced by an analogous procedure) that confers a Lac⁻ phenotype on mRZ352 (Reznikoff et al., 1982). After this heteroduplex was extended in vitro with DNA polymerase I and ligated, the mixture was extracted with phenol and ether and precipitated with ethanol. JI101 was transfected and Lac⁺ phage were screened for on indicator plates. Approximately 1 to 2% of the phage were reverted to a Lac⁺ phenotype. Incorporation of the lacZ mutation into mRZ352 was confirmed by DNA sequence analysis (90% of the phage that had incorporated the lacPUV5 promoter also contained the lacZ mutation).

(f) Recombination onto λplac5

The α-complementing fragment encoded on the M13 vector allows qualitative analysis of the effects of the mutations, but this fragment is unstable and β-galactosidase activity cannot be measured reliably (Bukhari & Zipser, 1973). Therefore, the mutations were recombined onto a λ derivative containing the entire lacZ gene. λplac5-T743 contains the lac control region and lacZ gene, but has a Lac⁻ phenotype due to the promoter mutation T743. A recombination event between the M13 vector and the λplac5-T743, which recombines in the lacPUV5 promoter and the restriction sites in the NH₂ terminus of the lacZ gene, will also introduce the mutations described above (these mutations are located between the promoter and the restriction sites; see Fig. 2).

λplac5-T743 (cl8578am7) was grown lytically on RZ411, which was chronically infected with the mutant mRZ361 (or mRZ352). After 3 h at 37°C, the resulting lysates were screened on XA103 for Lac⁺ λ phage. The supF strain allows growth of the λ phage, whereas the M13 derivatives require both supE and an F factor. DNA from the Lac⁻ λ phage was isolated as described by Maniatis et al. (1982) and digested with HindIII or EcoRI. Recombinant λ phage that contained the HindIII site but not the EcoRI site (see Fig. 2) were studied further. Single lysogens of the recombinant λplac5:361 or λplac5:352 phage were made in RZ201 and β-galactosidase was assayed at 32°C in M9 salts supplemented with 0.4% glucose and 0.2% Casamino acids according to Miller (1972).

(g) Computer methods

Most computer analyses were done using programs of the Delila system (Schneider et al., 1982). Two methods were used to predict secondary structures. In the “dot matrix” method, similar to that described by Maizel & Lenk (1981), all potential structures could be looked at. This was especially useful in noticing potential structural changes due to the sequence changes introduced by the mutations. The second method was a “dynamic programming” approach (program from E. Meyers), similar to that described by Zuker & Stiegler (1981), which finds the single best structure over some region (≤ 120 bases in our program) based on a set of rules. We always used the rules of Tinoco et al. (1973).
(h) Purification of β-galactosidase

β-Galactosidase was purified from a 100 ml culture of RZ201, lysogenic for the specified λ derivative, grown in superbroth (Davis et al., 1982) to an absorbance at 550 nm of 0.9 unit. Heat induction of the lysogen resulted in an approximately 200-fold overproduction of β-galactosidase after 3 h at 37°C. At this point the cells were collected by centrifugation and resuspended in 10 ml of buffer A (20 mM-Tris·HCl (pH 7.5), 5 mM-MgCl₂, 200 mM-NaCl, 5 mM-2-mercaptoethanol, 125 μg phenylmethylsulfonylfluoride). The suspension was freeze-thawed and then incubated for 10 min on ice with 2 mg lysozyme. The incubation was continued for 5 min at 37°C, followed by the addition of sodium dodecyl sulfate (0.05% final concentration). DNase I was added (1 mg) and the incubation continued for 5 min at 37°C. The cellular debris was removed by centrifugation at 30,000 g at 4°C. The crude extract was brought to 45% saturation with (NH₄)SO₄ and the precipitate was collected by centrifugation. The pellet was resuspended in 0.5 to 1 ml of buffer A and dialyzed for 1 to 2 h versus buffer A; 0.5 ml was loaded on a 5% to 20% sucrose gradient (11 ml preformed in buffer A) and centrifuged for 12 h at 240,000 g in a Beckman SW41 rotor. The active fractions were pooled and β-galactosidase was determined to be ~90% pure by sodium dodecyl sulfate/polyacrylamide gel electrophoresis (Laemmli, 1970).

(i) NH₂-terminal sequence analysis

The NH₂-terminal sequence analysis was done by the method described by Tarr (1982). The phenylthiohydantoin-amino acid (PTH) analysis was performed as described by Hunkapiller & Hood (1983).

3. Results

(a) Isolation of mutations

A vector was constructed to facilitate the screening and sequencing of mutations located within the lac regulatory region. This vector (mRZ361) is a derivative of M13mp8 and is shown in Figure 1. It contains an EcoRI restriction site located from the lacUV5 promoter and a series of unique restriction sites introduced into the NH₂ terminus of the lacZ gene. All the sequences required for

![Diagram](https://via.placeholder.com/150)

**Fig. 1.** Diagram of mRZ361. Restriction sites are indicated as follows: AII, AvaII; RI, EcoRI; SI, SalI; PI, PsiI; HIII, HindIII; BII, BglII. DNA from pBR322 corresponds to nucleotides 3726 to 4359 (Sutcliffe, 1979).
transcription and translation initiation are located within the 207 bp between the EcoRI and SalI restriction sites. Only a portion of the lacZ gene is contained on the vector (encoding the α-complementing fragment). The peptide synthesized from this fragment will complement a deletion in the NH2 terminus of the lacZ gene and thus, in the appropriate host strain, mRZ361 will confer a Lac⁺ phenotype.

mRZ361 was mutagenized by either growing the phage on a bacterial mutator strain or by treatment in vitro with hydroxylamine (see Materials and Methods). In order to study only those mutations located in the lac control region, the EcoRI-SalI restriction fragment was subcloned into an unmutagenized vector and the Lac phenotype was screened. Phage that conferred either an increase or a decrease in β-galactosidase activity were isolated and sequenced. The mutations obtained that are located between +1 and +70 (with respect to the beginning of the mRNA) are shown in Figure 3. All mutations are single point mutations; no double mutations were obtained.

(b) mRNA secondary structure

Mutations 32, 8, 6, 1 and 31 all increase β-galactosidase activity and are located near two potential initiation codons (mutation 8 actually changes the first AUG to AUU). Owing to their location and phenotype, it was thought that there might be an inhibitory secondary structure in the mRNA that these mutations disrupt. A relatively stable potential secondary structure was found in this region and is shown diagrammatically in Figure 4(a). All five point mutations would destabilize this structure. The arrow in Figure 4 indicates the beginning of the DNA (27 bp) introduced during the construction of M13mp8 and mRZ361 in order to create unique restriction sites in this region. Thus, this secondary structure is possible only in this construction and cannot be formed in the mRNAs encoded by the lacZ gene or M13mp8.
Further support for the existence of this secondary structure in vivo came with the isolation of pseudorevertants of mutations 76 and 2 (both of these mutations strongly decrease expression of the lacZ gene). mRZ361-76 and mRZ361-2 were grown on a bacterial mutator strain and phage that had increased levels of β-galactosidase activity were isolated. A number of these were sequenced, and Figure 4(b) shows the location of these second site mutations. The fact that both of these mutations give rise to pseudorevertants located in the sequence encoding the potential stem-and-loop structure, suggests that regardless of the mechanism by which these mutations decrease β-galactosidase activity, the most common way to increase lacZ expression is by disrupting this secondary structure. This can
be explained by the observation that there are many sites that can be mutated leading to a sequence change that destabilizes the secondary structure, while presumably only a few sites can be mutated to revert the original defect (this is probably true for 76, but see below for 2). Pseudorevertants 76:50, 2:52 and 2:53 are located outside the region in which the ribosome makes direct contact with the message (determined by RNase protection experiments: Maizels, 1974), thus it is unlikely that these mutations affect contact sites with the ribosome during initiation.

(c) Deletion of the inhibitory secondary structure

To test further the hypothesis that the secondary structure shown in Figure 4 exists in vivo and to study the mutations obtained in the absence of this structure, a 9 bp deletion was made in mRZ361. This deletion, which removes nucleotides from +63 to +71 (with respect to the transcriptional start site, see Fig. 4), should prevent the formation of the stem-and-loop. It in fact results in a substantial increase in β-galactosidase activity. Since the ribosome binding site extends to approximately +55, the 9 bp deletion is located outside the region contacted by the ribosome during initiation.

It is assumed that the amino acid deletion does not alter the specific activity of β-galactosidase because amino acid changes and deletions in this region have previously been shown to have no effect on the enzyme’s specific activity (Sarthy et al., 1979).

mRZ361 with the 9 bp deleted was named mRZ352 and most of the mutations isolated were transferred into this vector (see Materials and Methods). Quantitation of the effect of the mutations was determined by measuring β-galactosidase activity from lysogens of λpZac5 derivatives containing the mutations (see Materials and Methods and Fig. 2).

(d) Mutations affecting mRNA secondary structure

Mutations 31, 1, 6 and 32 were proposed to increase lacZ expression by destabilizing the inhibitory stem-and-loop. Table 2 shows the β-galactosidase activity of these mutations in λpZac5:361 and λpZac5:352 (mutation 8 is discussed in section (h), below). Comparing the wild-type λpZac5 : 361 to λpZac5 : 352, it can be seen that β-galactosidase activity increases 5-8-fold. Mutations 1 and 31 in λpZac5:361 also result in an approximately 5-8-fold increase in β-galactosidase activity. These single point mutations, therefore, destabilize the secondary structure and increase lacZ expression to the same degree as the 9 bp deletion. The calculated free energy for the stem-and-loop in 361-1 or 361-31 is -5.2 kcal/mol (1 cal = 4.184 J) as compared to -12.4 kcal/mol for 361 (calculated according to Tinoco et al., 1973). Mutations 1 and 31 in conjunction with the 9 bp deletion (in λpZac5:352) express the lacZ gene at a level very close to the wild-type λpZac5:352. This supports the idea that these mutations increase β-galactosidase activity by destabilization of the secondary structure and when this is not present (due to the 9 bp deletion), their effect is negligible. These results
TABLE 2

$\beta$-Galactosidase activity of mutations that affect mRNA secondary structure

<table>
<thead>
<tr>
<th>Mutation</th>
<th>$\lambda$plac5:361 % wt (361)</th>
<th>$\lambda$plac5:352 % wt (361)</th>
<th>$\lambda$plac5:352 % wt (352)</th>
</tr>
</thead>
<tbody>
<tr>
<td>wt</td>
<td>100</td>
<td>581</td>
<td>100</td>
</tr>
<tr>
<td>31</td>
<td>582</td>
<td>628</td>
<td>107</td>
</tr>
<tr>
<td>I</td>
<td>530</td>
<td>569</td>
<td>98</td>
</tr>
<tr>
<td>6</td>
<td>383</td>
<td>380</td>
<td>65</td>
</tr>
<tr>
<td>32</td>
<td>197</td>
<td>482</td>
<td>83</td>
</tr>
<tr>
<td>33</td>
<td>639</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>8.1</td>
<td>85</td>
<td>15</td>
</tr>
<tr>
<td>2:31</td>
<td>87</td>
<td>152</td>
<td>26</td>
</tr>
<tr>
<td>37</td>
<td>3.9</td>
<td>410</td>
<td>70</td>
</tr>
</tbody>
</table>

The values shown are averages from 3 to 6 independent sets of assays. Each assay was done in duplicate and normalized with respect to $\lambda$plac5:361. The relative error is less than 10%. $\lambda$plac5:361 produced 520 units of $\beta$-galactosidase activity (Miller, 1972). wt, wild type.

† $\lambda$plac5:352 contains a 9 bp deletion that removes sequences required for the formation of the proposed secondary structure.

‡ The values shown are for mutations in $\lambda$plac5:352, normalized with respect to $\lambda$plac5:361.

§ The values shown were normalized with respect to the average value of $\lambda$plac5:352.

also suggest that at these positions in the ribosome binding site, C to U transitions do not affect the interaction of the mRNA with the ribosome.

Mutation 6, however, only increases lacZ expression 3.8-fold in $\lambda$plac5:361. When this mutation is examined in $\lambda$plac5:352, it actually decreases $\beta$-galactosidase activity compared to wild-type $\lambda$plac5:352. In fact, the $\beta$-galactosidase activity is identical either with or without the 9 bp deletion (in $\lambda$plac5:361 or $\lambda$plac5:352). This indicates that mutation 6 can destabilize the stem to the same extent as the 9 bp deletion (otherwise the $\beta$-galactosidase activity should be greater in $\lambda$plac5:352-6 than in $\lambda$plac5:361-6), but that the change introduced by this mutation also has a negative effect on translation initiation.

Mutation 32 is located in the base of the stem-and-loop structure shown in Figure 4, and should destabilize it slightly (from a $\Delta G = -12.4$ kcal/mol to a $\Delta G = -10.2$ kcal/mol). This structure would then have a greater probability of opening, which may account for the twofold increase $\beta$-galactosidase activity. However, in conjunction with the 9 bp deletion, mutation 32 causes a decrease in synthesis of $\beta$-galactosidase by about 17% (compared to wild-type $\lambda$plac5:352). This inhibition could be explained by a new structure in 352-32 mRNA, involving G-A-A-A-U at +30 to +34, and A-U-U-U-C at +19 to +23. This structure would involve the Shine–Dalgarno sequence and could be sufficient to explain the decrease in lacZ expression.

(e) Alternative secondary structures

We believe that three mutations, 33, 37 and 2, may alter lacZ expression by changing the stability of alternative secondary structures in the mRNA. These
structures are shown in Figure 5 and the β-galactosidase activity for these mutations is reported in Table 2. Mutation 33 decreases the free energy of formation of a competing secondary structure (Fig. 5(a)), which when formed would prevent the formation of the inhibitory stem-and-loop shown in Figure 4. This could then explain the sixfold increase in β-galactosidase activity manifested by \( \lambda \text{plac}5:361-33 \). Mutation 37 increases the stability of a structure that involves the Shine–Dalgarno sequence as well as the initiation codon (Fig. 5(b)). It is possible that formation of this structure leads to the 26-fold reduction in lacZ expression (Table 2). In conjunction with the 9 bp deletion, a slight inhibition is still observed. This may be due to the introduction of a G residue near the Shine–Dalgarno sequence or the presence of a weak structure 5’ to the Shine–Dalgarno sequence that is stabilised by mutation 37 (not shown). Figure 5(c) shows a potential structure that is stabilised in 361-2 mRNA and Figure 5(d) shows diagrammatically a structure that can form in 352-2 mRNA. Pseudorevertant 2:31 is also shown in these diagrams. The estimated strength of these structures corresponds roughly to the inhibitory effect seen on lacZ expression.

Mutations 34 and 7 both increase lacZ expression approximately threefold, and mutation 9 decreases expression fourfold (data not shown). These mutations are located outside the ribosome binding site and cannot be readily explained by simple, stable secondary structures. However, structures whose \( \Delta G = -2 \text{ kcal/mol} \) can be drawn in the mRNA that precedes the Shine–Dalgarno sequence and mutations 34 and 7 decrease the stabilities of these structures. Thus these structures may account for the observed mutant phenotypes.

Mutation 77 has no effect on this fully induced level of lac expression in the \( \lambda \text{plac}5:361 \) construction and was instead presumably isolated due to its 0° phenotype.

(f) Mutations affecting the Shine–Dalgarno sequence

Two mutations were isolated that affect the sequence complementary to the 3’ end of 16S rRNA (3’ A-U-U-C-C-U-C-C-A). Mutation 76 decreases the four-base complementarity existing in the lacZ mRNA and mutation 11 increases the complementarity to five bases (see Fig. 3). Table 3 shows the effect mutations 76 and 11 have on lacZ expression in \( \lambda \text{plac}5:361 \) and \( \lambda \text{plac}5:352 \). In the presence of the inhibitory secondary structure (\( \lambda \text{plac}5:361 \)), decreasing the length of the Shine–Dalgarno sequence (mutation 76) reduces β-galactosidase activity 12-fold. Lengthening the Shine–Dalgarno sequence (mutation 11) increases lacZ expression 2.8-fold. Thus, the extent of the complementary sequence has a strong effect on translation initiation.

As was noted previously, the 9 bp deletion increases β-galactosidase activity 5.8-fold (compare wild-type \( \lambda \text{plac}5:361 \) to \( \lambda \text{plac}5:352 \)). Mutation 76 in this background (\( \lambda \text{plac}5:352-76 \)) only reduces β-galactosidase activity sixfold and mutation 11 has no effect. The presence of mutation 11 on the \( \lambda \text{plac}5:352 \) was checked by growing mRZ352-T743 on a \( \lambda \text{plac}5:352-11 \) lysogen, and isolating Lac+ recombinant M13 phage. T743 is a promoter mutation that confers a Lac− phenotype on the M13 vector, mRZ352; recombination of the promoter from the
Fig. 5. Potential secondary structures stabilized by mutations 33, 37 and 2. (a) A secondary structure that competes with the inhibitory structure is shown. Mutation 33 decreases the free energy of formation of this competing structure from $-10.2$ kcal/mol to $-20.2$ kcal/mol. To see the complete extent of competition, the structure shown here should be compared with the structure in Fig. 4. (b) Mutation 37 increases the stability of the structure shown from $\Delta G = -12.4$ kcal/mol to $\Delta G = -19.2$ kcal/mol. (c) Mutation 2 in 361 mRNA allows the formation of the structure shown, increasing its stability by approximately $-10$ kcal/mol. b. bases. (d) In 352 mRNA mutation 2 increases the stability of the structure shown by approximately $-10$ kcal/mol. b. bases.
**Ribosome Binding Site Mutations**

![Diagram of mRNA secondary structure](image)

**Fig. 6.** A potential mRNA secondary structure encoded by $\lambda$plac5:361-5. The structure is similar to that shown in Fig. 4 but is modified to accommodate the AUG $\rightarrow$ GUG sequence change.

$\lambda$plac5 derivative reverts this to Lac$. Two Lac$ M13 phage were isolated and the DNA near the beginning of the $\lambda$acZ gene was sequenced. Both phage contained mutation 11, confirming that the original $\lambda$plac5:352-11 also contained this mutation. The fact that this mutation did not increase lacZ expression in $\lambda$plac5:352 indicates that increasing the length of the Shine–Dalgarno sequence will not necessarily increase translation initiation.

**(g) NH$_2$-terminal sequence of $\beta$-galactosidase**

There are two potential initiation codons for the lacZ gene and three mutations were isolated that are located in the first initiation codon (8, AUA; 3, ACG; 5, GUG: see Fig. 3). Two pseudorevertants were also isolated that contained sequence changes in the first initiation codon (3:55, ACU and 5:8, GUA: pseudorevertants of mutations 3 and 5, respectively, which were isolated as described previously). In order to determine which initiation codon was being used in each of these mutants, the mutations were recombined onto $\lambda$plac5 and $\beta$-galactosidase was isolated from strains lysogenic for these derivatives (see Materials and Methods). The NH$_2$-terminal sequence was determined by manual Edman degradation and the results are shown in Table 4. A total of 79% of the $\beta$-galactosidase synthesized from $\lambda$plac5 (which is identical to the chromosomal lacZ gene) had Thr-Met-Ile-Thr as the NH$_2$-terminal sequence. This is the correct sequence if initiation occurs at the first initiation codon and the Met is removed by aminopeptidase; 21% of the protein contained the residues Met-Ile-Thr at the NH$_2$ terminus indicating that initiation was occurring at the second initiation codon. In this case the Met was not cleaved by aminopeptidase. The percentage of each protein was determined by comparing the amount of Met versus Ile released during the second cycle of Edman degradations. The first cycle was not measured, due to the inaccuracy in quantitating Thr. These results differ slightly from
Previous results in which Thr was identified as the sole NH$_2$ terminus of \( \beta \)-galactosidase (Brown et al., 1966). It is possible that previous studies may have missed the 10 to 20\% initiation occurring at the second initiation codon (I. Zabin, personal communication), or that the difference in strains and growth procedures could account for this discrepancy.

A lysogen of \( \lambda p l a c 5 : 361 \) (which contains an additional 27 bp inserted into the NH$_2$ terminus of the lacZ gene) directed the synthesis of \( \beta \)-galactosidase proteins that used the first initiation codon 43\% of the time and the second initiation codon 57\% of the time. This result may be explained by the fact that both initiation codons are involved in the secondary structure and that the frequency at which the initiation codon is used reflects the percentage of the time it is “melted out”. When this secondary structure is deleted (\( \lambda p l a c 5 : 352 \)), the first initiation codon is again the one that is preferred (65\% of the protein initiates at the first AUG). In the absence of secondary structure, mutation 5 (GUG; \( \lambda p l a c 5 : 352-5 \)) directs the synthesis of proteins that initiate translation 28\% of the time at the first initiation codon and 72\% of the time at the second initiation codon. Both pseudorevertants 5:8 (GUA) and 3:55 (ACU) direct the synthesis of \( \beta \)-galactosidase that initiates only from the second initiation codon (thus the GUA

### Table 3

<table>
<thead>
<tr>
<th>Mutation</th>
<th>( \lambda p l a c 5 : 361 ) ( % ) wt (361)</th>
<th>( \lambda p l a c 5 : 352 ) ( % ) wt (352)</th>
</tr>
</thead>
<tbody>
<tr>
<td>wt</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>76</td>
<td>79</td>
<td>96</td>
</tr>
<tr>
<td>11</td>
<td>278</td>
<td>551</td>
</tr>
</tbody>
</table>

The values shown were determined as described in the legend to Table 2.

### Table 4

Frequency of initiation codon usage

<table>
<thead>
<tr>
<th>( \lambda ) Derivative</th>
<th>Sequence</th>
<th>1st initiation codon</th>
<th>2nd initiation codon</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \lambda p l a c 5 )</td>
<td>AUG</td>
<td>Expt 1</td>
<td>Expt 2</td>
</tr>
<tr>
<td>( \lambda p l a c 5 : 361 )</td>
<td>AUG</td>
<td>79</td>
<td>79</td>
</tr>
<tr>
<td>( \lambda p l a c 5 : 352 )</td>
<td>AUG</td>
<td>69</td>
<td>65</td>
</tr>
<tr>
<td>( \lambda p l a c 5 : 352-5 )</td>
<td>GUG</td>
<td>34</td>
<td>28</td>
</tr>
<tr>
<td>( \lambda p l a c 5 : 361-5:8 )</td>
<td>GUA</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( \lambda p l a c 5 : 361-3:55 )</td>
<td>ACU</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

The values shown are the percentages of total initiation events.
† The sequence changes all occur in the first initiation codon.
‡ Each experiment represents the sequence results from an independent \( \beta \)-galactosidase preparation.
and ACU do not direct initiation of translation). In both 5:8 and 3:55 the β-galactosidase NH₂ terminus was Met-Ile-Thr, substantiating the observation that when initiation occurs at the second AUG the Met residue is not removed by aminopeptidase. β-Galactosidase was not sequenced from strains containing mutation 3 (ACG) or 8 (AUA).

(h) Mutations located in the first initiation codon

The effect of mutations 8, 3 and 5, and pseudorevertants 5:8 and 3:55, on translation initiation was examined by measuring β-galactosidase activity from lysogens of the λlac5 derivatives (Table 5). In mRNA transcribed from λlac5:361, three mutations weakened the inhibitory secondary structure shown in Figure 4 (5:8, 8 and 3:55). In each case the third position of the AUG was changed (to either an A or a U). In mutation 8 (AUA) and pseudorevertant 3:55 (ACU), β-galactosidase activity was increased approximately threefold. In 5:8 (GUA), β-galactosidase activity only increased 15% above wild-type λlac5:361 (Table 5). Analysis of potential secondary structures revealed that 5:8 creates the sequence C-II-G-U (+37 to +40), which can pair with A-C-A-G (+26 to +29). This new structure involves the initiation codon base-pairing with a portion of the Shine-Dalgarno sequence, and would be expected to inhibit translation initiation.

Introduction of the 9 bp deletion results in only a 10% increase in β-galactosidase activity for 5:8, 8 and 3:55 (Table 5). This supports the hypothesis that the structure shown in Figure 4 is no longer inhibitory to lacZ expression when the third position of the AUG codon is changed. Mutations 8 (AUA) and 3:55 (ACU) allow lacZ expression at 60% of the level of wild-type λlac5:352. 3:55 was shown above to direct the synthesis of β-galactosidase that initiates only at the second initiation codon. These results indicate that the second initiation codon can be used quite efficiently, even though the distance from the Shine–Dalgarno sequence is increased from 7 bases (first AUG) to 13 bases (second AUG). Mutation 5:8 also produces β-galactosidase that is initiated only at the

<table>
<thead>
<tr>
<th>Mutation</th>
<th>Sequence†</th>
<th>λlac5:361</th>
<th>λlac5:352</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>% wt (361)</td>
<td>% wt (352)</td>
<td>% wt (352)</td>
</tr>
<tr>
<td>wt</td>
<td>AUG</td>
<td>100</td>
<td>581</td>
</tr>
<tr>
<td>5</td>
<td>GUG</td>
<td>25</td>
<td>145</td>
</tr>
<tr>
<td>5:8</td>
<td>GUA</td>
<td>116</td>
<td>123</td>
</tr>
<tr>
<td>x</td>
<td>AUA</td>
<td>304</td>
<td>330</td>
</tr>
<tr>
<td>3</td>
<td>ACG</td>
<td>3.5</td>
<td>112</td>
</tr>
<tr>
<td>3:55</td>
<td>ACU</td>
<td>310</td>
<td>348</td>
</tr>
</tbody>
</table>

The values were obtained as described for Table 2.

† All sequence changes are in the first initiation codon.
second AUG, but at a reduced efficiency possible due to the new structure described above.

Mutation 5 changes the first AUG to GUG and reduces lacZ expression fourfold in both λpλac5:361 and λpλac5:352. This was somewhat surprising since the second AUG can allow lacZ expression at 60% the wild-type level (see above). A modified version of the structure shown in Figure 4 is stabilized by mutation 5 ($\Delta G = -14.6$ kcal/mol when the lower portion is reorganized as shown in Fig. 6), which may explain the lower level of β-galactosidase activity in λpλac5:361. In λpλac5:352, the structure described previously for 5:8 is strengthened by $-2.1$ kcal/mol (C-U-G-U-G (+37 to +41) with C-A-C-A-G (+25 to +29)), and so should be more inhibitory.

Table 6 shows a breakdown of β-galactosidase units according to the initiation codon used in λpλac5:352. There is a tenfold reduction in protein synthesized from the first initiation codon when it is a GUG (5) compared to AUG (wild-type). The second AUG is also used at a reduced frequency in both 5 and 5:8. When the first initiation codon is changed to ACU, there is a twofold increase in initiation at the second AUG compared to wild-type. The levels of β-galactosidase activity may reflect the mRNA secondary structure in these messages, although it is also possible that the effects are caused by changing the nucleotides located between the Shine–Dalgarno sequence and the initiation codon.

The decrease in β-galactosidase activity caused by mutation 3 (ACG) is larger than would be predicted by either the loss of the first initiation codon or the slight increase in stability of the structure shown in Figure 4 ($-1$ kcal/mol). This mutation also maintains a G residue four bases before the second AUG. A statistical analysis has shown that G residues are discriminated against at this position, and may be inhibitory to ribosome binding (Stormo et al., 1982). In λpλac5:352, a new structure may form between the A-C-G-A-C and its complement located 40 bases downstream, which may contribute to the fivefold decrease in β-galactosidase activity.

**Table 6**

| Units of β-galactosidase activity corresponding to initiation codon usage |
|---|---|
| **Mutation** | **Sequence** | **β-Galactosidase units** |
|  |  | **1st init. codon** | **2nd init. codon** |
| wt | AUG | 1950 | 1050 |
| 5 | GUG | 210 | 540 |
| 5:8 | GUA | 0 | 630 |
| 3.55 | ACU | 0 | 1800 |

† The activity of the mutations is based on the values obtained in λpλac5:352. Mutations 5:8 and 3.55 are assumed to use only the second initiation codon in λpλac5:352 (since they used only the second initiation codon in λpλac5:361).

†† Units of β-galactosidase activity are based on the average wild-type value λpλac5:352 (3000 units), the percentage of activity of each of the mutations and the percentage of time for which the particular initiation is used.
4. Discussion

The isolation of 16 point mutations near the beginning of the lacZ gene has allowed the analysis of some of the sequence determinants defining this ribosome binding site and has indicated the presence of an inhibitory secondary structure. This secondary structure (Fig. 4) is a fortuitous result of the construction of the M13 cloning vector (mRZ361), which allowed the isolation of several mutations that would not have been obtained otherwise.

(a) mRNA secondary structure

Several lines of evidence were presented that support the existence in vivo of the inhibitory secondary structure shown in Figure 4. Two single point mutations that disrupt this structure (1 and 31) increase β-galactosidase activity to the same extent as a 9 bp deletion that prevents its formation (5.8-fold increase in lacZ expression). When these point mutations were combined with the 9 bp deletion, no significant increase in activity was seen. Furthermore, seven pseudorevertants of either mutation 76 or 2 (which have low levels of β-galactosidase activity) are located in this structure; three of these are located outside the ribosome binding site.

Three other mutations (37, 2 and 33) can also be explained by secondary structure considerations. Mutation 33 increases lacZ expression sixfold and can be explained by the stabilization of a competitive secondary structure. This structure, shown in Figure 5, involves some of the same bases as the inhibitory secondary structure described previously; thus they are mutually exclusive. Mutation 37 stabilizes an extension of the stem-and-loop to involve the Shine-Dalgarno sequence (Fig. 5). This mutation decreases lacZ expression 26-fold and indicates the importance of an available Shine-Dalgarno sequence for translation initiation. Mutation 2, like 37, stabilizes the inhibitory structure and decreases β-galactosidase synthesis. Much of the structure that exists in 361-2 mRNA also exists in 352-2 mRNA, and so expression is reduced in λlacs:352-2 as well. All of these mutations demonstrate the inhibitory effect on translation initiation of having the initiation codon involved in a stable secondary structure.

Although the direct effect of mRNA structure on ribosome binding can adequately explain the phenotype of most of these mutations, it is also possible that the mRNA half-life is altered (the mRNA half-life was not measured in this study). Cannistraro & Kennell (1979) have determined the half-life of mRNA transcribed from several lac0° mutants and found that it varied over a less than twofold range. Since the mutations described here result in a variation in β-galactosidase activity of 180-fold, the effect on mRNA half-life would have to be much larger than that of the 0° mutations to account for these results.

(b) Second codon

Three mutations that were isolated because they disrupt the inhibitory secondary structure, are located in the second codon. These mutations change this codon from ACC to CCC, AUC or ACU (mutations 6, 1 and 31, respectively). The
two C to U transitions have no effect on lacZ expression in the absence of the inhibitory secondary structure. However, the A to C transversion (mutation 6) decreases β-galactosidase activity by 35%.

This mutation creates a codon that is rarely used in *Escherichia coli*, which could cause translation to be stalled while the minor tRNA species diffuses into the ribosome (Konigsberg & Godson, 1983; Ikemura, 1981). Alternatively, there may be a differential interaction of the initiator tRNA with the incoming tRNA, as proposed by Bossi & Roth (1980) to explain the efficiency of suppressor tRNAs (Bossi, 1983). Finally, an extended interaction between the anticodon loop of the initiator tRNA and the initiation codon (involving the U33 residue in the tRNA and an A following the AUG) may be required for efficient initiation, and mutation 6 would prevent this interaction from occurring (Taniguchi & Wiessmann, 1978).

(c) Shine–Dalgarno sequence

It has not yet been possible to demonstrate a clear correlation between the strength of a ribosome binding site (in terms of its ability to promote high levels of translation) and the length of the sequence that is complementary to the 16S rRNA. We have presented data that suggest that the same Shine–Dalgarno sequence will affect translation initiation differently depending upon the mRNA secondary structure. In accordance with the models presented earlier, the mRNA, which can form the inhibitory secondary structure, may bind to the 30 S subunit by virtue of its Shine–Dalgarno sequence and stimulate binding of the fMet-tRNA. If the interaction of the anticodon loop of the fMet-tRNA with the initiation codon must await the opening of the secondary structure, the stability of this pre-initiation complex may be largely determined by the extent of the complementarity with the 16S rRNA. We have shown that a mutation that shortened the Shine–Dalgarno sequence resulted in a 12-fold reduction in β-galactosidase activity. Likewise, a mutation that lengthened the Shine–Dalgarno sequence from four to five bases (a ΔG change of −2.2 kcal/mol) resulted in a 2.8-fold increase in lacZ expression.

When these same mutations were examined in conjunction with a 9 bp deletion that prevented the formation of the inhibitory secondary structure, quite different results were seen. In this case, the mutation that decreased the length of the Shine–Dalgarno sequence only reduced lacZ expression sixfold. The mutation that increased the length of the Shine–Dalgarno sequence had no effect. We interpret this to mean that the rate of the association of the fMet-tRNA with the initiation codon was faster than the dissociation of the mRNA from the pre-initiation complex when the Shine–Dalgarno sequence was four bases long. Increasing this sequence to five bases would then make no difference, but the Shine–Dalgarno sequence could still be decreased to the point where dissociation occurred more rapidly than initiation. One would predict that particularly long Shine–Dalgarno sequences would enhance translation initiation only if the initiation codon was not available (due to secondary structure) or if it was a particularly poor one (GUG or UUG). The distance between these two sequences and possibly the particular nucleotides in the region may also affect this step (see below).
We have determined the NH₂-terminal sequence of β-galactosidase isolated from induced λ lysogens and found that both initiation codons can be used (see Fig. 3 and Table 4). This was somewhat surprising in that the initiation codons are 7 and 13 bases away from the Shine–Dalgarno sequence, which mutation 76 strongly implicates as the only Shine–Dalgarno sequence in the region. In λplac5 (which contains the completely normal lacZ gene), initiation occurs approximately 80% of the time at the first AUG and 20% of the time at the second AUG. This indicates that even in the wild-type gene some initiation occurs at the second AUG which has a 13-base spacer region.

Introduction of the secondary structure that involves both initiation codons (λplac5:361) alters the pattern of initiation so that both codons are used almost equally. We believe this may reflect the percentage of time that each initiation codon is released from the secondary structure. Deletion of the secondary structure (λplac5:358) shifts the initiation pattern so that the first initiation codon is again preferred. The frequency at which the second AUG is used is slightly higher for λplac5:352 than λplac5, but this may reflect some secondary structure in this message that has not yet been identified (there are 18 bases in λplac5:352 message that are not in λplac5 message).

Two mutations that alter the first AUG (to either AUA or ACU) decrease β-galactosidase activity to approximately 60%. NH₂-terminal sequence data for 3:55 (ACU) indicate that all initiation is occurring at the secondary structure. This suggests that changing the spacing from 7 to 13 bases decreases expression less than twofold. From this result we conclude that the spacing between the Shine–Dalgarno sequence and the initiation codon, within a certain range, is not necessarily a major factor in determining the efficiency of translation initiation.

Mutations 5:8 and 3 differ from those described above by one base (GUA versus AUA and ACG versus ACU, respectively), and in both cases initiation occurs at the second initiation codon. However, these mutations reduce β-galactosidase activity fivefold (compared to less than twofold for the other mutations). The differences in the sequences in both cases involves the introduction of a G residue, in either the first or third position of the first initiation codon. These results suggest that the sequences located between the Shine–Dalgarno sequence and the initiation codon have a large effect on translation initiation and may alter the efficiency at which an initiation codon can be used. It is difficult to separate effects due to the primary sequence versus creation of new secondary or tertiary structures, but in both of these cases secondary structures were found that may account for the lower β-galactosidase activity.

In the bacteriophage T7 0-3 gene two initiation codons also exist, located 8 and 14 bases from the Shine–Dalgarno sequence. In this gene a mutation changing the first AUG to ACG decreased translation initiation approximately tenfold (Dunn et al., 1978). Though the spacer region is only one base longer than in the lacZ gene, the complementarity to the 16 S rRNA is displaced three bases away from the 3’ end of the 16 S rRNA. It is possible that this may shorten the length of the spacer region that allows for efficient translation initiation. mRNA secondary structures were not reported for this system.
A mutation that changed the first AUG to GUG both decreased β-galactosidase activity and altered the frequency of initiation at this codon (Table 6). Calculations of the amount of β-galactosidase initiating at the GUG revealed a tenfold decrease compared to wild-type. In the bacteriophage T4 rIIIB gene a mutation was also obtained that changed the initiation codon from AUG to GUG, but this mutation only reduced translation initiation threefold (L. Gold, personal communication). We believe that the difference between the lacZ and T4 rIIIB genes can be explained by creation of a new secondary structure in the lacZ mRNA. This structure involves the GUG and part of the Shine–Dalgarno sequence and would be expected to inhibit initiation strongly at the GUG and weakly at the second initiation codon (AUG).

(e) Conclusions

From these studies we can conclude that mRNA secondary structure in the ribosome binding site can drastically affect translation initiation. We have also shown that the length of the Shine–Dalgarno sequence can influence the rate of translation initiation, but the magnitude of the effect is probably dependent on the rate-limiting step. We predict that for an unstructured message with a favorably spaced AUG, increasing the length of a four to five-base Shine–Dalgarno sequence will probably not affect translation initiation. In addition, it was found that two initiation codons in a single ribosome binding site can be used at equal frequencies and that for lacZ, increasing the spacer region from 7 to 13 bases does not drastically affect the efficiency of initiation. Finally, sequence changes between the Shine–Dalgarno sequence and the initiation codon can affect the efficiency of translation initiation. In particular, introduction of additional G residues always resulted in decreased lacZ expression. This may be due to direct inhibition by G residues on ribosome binding, increased secondary structure, or both.

We thank R. Johnson for critical reading of the manuscript. This work was supported by grant GM-19670 from the National Institutes of Health and a grant from the 3M Foundation. G.D.S. was supported by grant GM-28755 from the National Institutes of Health and R.L.N. was supported by NIH program project grant CA23076 and NIH grant CA97175 to the McArdle Laboratory. The protein sequence data in this report were the initial contribution of the University of Wisconsin protein sequence–DNA synthesis facility.

REFERENCES


*Edited by M. Gottesman*