Volume 12 Number 9 1984 Nucleic Acids Research

# Selection of initiation sites by eucaryotic ribosomes: effect of inserting AUG triplets upstream from the coding sequence for preproinsulin

Marilyn Kozak

Department of Biological Sciences, University of Pittsburgh, Pittsburgh, PA 15260, USA

Received 14 December 1983; Revised and Accepted 5 April 1984

ABSTRACT

Recombinant plasmids that direct synthesis of rat preproinsulin under the direction of the SV40 early promoter have been used to probe the mechanism of initiation of translation. Insertion of an upstream AUG triplet that was out-of-frame with respect to the coding sequence for preproinsulin reduced the yield of proinsulin, in keeping with the predictions of the scanning model. The extent to which an upstream AUG codon interfered depended on sequences surrounding the AUG triplet; with two constructs (p255/20 and C2) the 5'-proximal AUG codon constituted an absolute barrier: there was no initiation at the downstream start site for preproinsulin. With two other constructs (p255/9, p255/21), however, proinsulin was made despite the presence of an upstream, out-of-frame AUG codon in a favorable context for initiation. In those cases the reading frame set by the first AUG triplet was short, terminating before the start of the preproinsulin coding sequence. The interpretation that ribosomes initiate at the first AUG, terminate, and then *reinitiate* at the AUG that directly precedes the preproinsulin coding sequence was tested by introducing a point mutation that eliminated the terminator codon: the resulting mutant made no proinsulin.

## INTRODUCTION

A growing body of evidence supports the idea that eucaryotic ribosomes might scan the 5'-end of messenger RNA, searching for an AUG triplet that can serve as the start site for translation (1-3). The scanning mechanism rationalizes the observation that translation begins at the 5'-proximal AUG triplet in most (>90%) eucaryotic mRNAs. The existence of some well-characterized mRNAs that contradict the first-AUG-rule, however, provoked the formulation of a modified model which takes into account both the position of a potential initiator codon (i.e., its proximity to the 5'-terminus) and the sequence context in which it occurs. The modified scanning model (4, 5) states that a 40S ribosomal subunit binds at the 5'-end of the message and migrates until it encounters the first AUG triplet: if the first AUG codon occurs in a favorable sequence context, all 40S subunits stop there and that AUG triplet uniquely serves as the initiator codon. If the first AUG triplet occurs in a suboptimal sequence context, however, some 40S subunits bypass that site and initiate

instead at another AUG codon that lies farther downstream. Based on a survey of published mRNA sequences, CC$\frac{A}{G}$CCAUGG has been tentatively identified as a favorable context for initiation by eucaryotic ribosomes (4,6). The prediction that adenosine in position -3 (i.e., 3 nucleotides upstream from the AUG) contributes in an important way to the ability of an AUG triplet to serve as an initiator codon has recently been confirmed by introducing point mutations in that position in a cloned preproinsulin gene (7). Those mutagenesis studies constitute a first step toward experimentally defining the "optimal sequence context" postulated by the modified scanning model. The experiments in the present report address two additional questions relating to the model: (i) Does insertion of an upstream AUG triplet diminish the production of (pre)proinsulin initiated from the normal start site that lies downstream? and (ii) Does the efficiency of an upstream AUG "barrier" vary as a function of sequence context? In addition to answering those questions, the experiments that follow revealed unexpectedly that eucaryotic ribosomes appear to be able to reinitiate after encountering a translational stop codon.

## EXPERIMENTAL PROCEDURES

### General Methodology

E. coli strain MM294 was routinely used as the host for transformation, which was carried out according to standard procedures (3, 8). When it was necessary to obtain plasmid DNA that could be cut by endonuclease BclI, strain MM294 was unsuitable because it methylates an A residue within the BclI recognition site. Therefore, transformation was carried out using E. coli SR19, which lacks the A-specific methylase. The restriction enzymes used for screening and restructuring plasmid DNA were used under the conditions recommended by the supplier, New England BioLabs. The chemical cleavage method of Maxam and Gilbert (9) was used for sequencing DNA. Plasmid DNA for transfection of COS-1 cells was purified by chromatography on BioGel A50 followed by banding in ethidium bromide/CsCl gradients (3).

Slightly subconfluent monolayers of COS-1 cells (10) were transfected with a mixture of plasmid DNA and high molecular weight carrier DNA exactly as described previously (3) except that the volume of DNA/$CaPO_4$ precipitate was reduced to 0.5 ml per 60 mm plate. Forty-eight hr after transfection the cells were labeled with $^{35}$S-cysteine, after which proteins were extracted, immunoprecipitated with antiserum against bovine insulin (Miles) and subjected to SDS-polyacrylamide gel electrophoresis as described (3).

RNA was prepared from the cytoplasmic fraction of transfected cells according to procedures described by Murphy et al. (11). Two-fold serial dilutions of de-

natured RNA were applied to a Gene Screen hybridization membrane (New England Nuclear) using a Hybri·dot apparatus (BRL). Hybridization was carried out with a heat-denatured $^{32}$P-labeled probe obtained by nick-translation of the BamHI-EcoRI fragment of p255. No significant differences in cytoplasmic mRNA levels were observed between wild type p255 and p255/2A, 15A and 16A. There was also no detectable variation in mRNA levels between p255/20 and p255/21, or between mutants C1 and C2.

Genesis of Plasmids p255/2A, p255/15A and p255/16A

The first step in the construction was to insert into p255 an 11 bp oligonucleotide, generating a derivative called p255/2. To accomplish this, p255 was linearized by digesting with HindIII, treated with alkaline phosphatase to remove the terminal phosphate moieties, and then annealed with the synthetic 5'-phosphorylated oligonucleotide 5'-AGCTTATGATC-3' / 3'-ATACTAGTCGA-5'. The oligonucleotide was custom synthesized by PL Biochemicals, Inc. Following incubation with T4 DNA ligase for 24 hr at 4°C, an aliquot of the reaction mixture was used to transform E. coli. Mutant p255/2, which had acquired one copy of the 11 bp insert, was initially identified by the presence of a cleavage site for BclI (T/GATCA). DNA sequence analysis subsequently confirmed the presence and orientation of the insert.

Mutant p255/2A was obtained next by inserting, at the BclI site of p255/2, the 14 bp oligonucleotide 5'-GATCGAAAGCTTTC-3'. (The objective here was merely to insert a sequence that would shift the upstream AUG triplet in p255/2 into a different reading frame from the preproinsulin coding sequence. The particular 14 bp oligonucleotide that I used for this purpose had been designed and used for another experiment; because it was on hand and fulfilled the need, I used it again here.) When the aforementioned oligonucleotide is self-annealed, it forms a partially base-paired duplex with 5'-protruding termini that are complementary to the overhangs generated by cutting p255/2 with BclI. The presence of a single copy of the 14 bp insert in p255/2A was confirmed by DNA sequence analysis.

A targeted mutagenesis procedure was used to obtain derivatives p255/15 and p255/16, which have an 11 bp insertion similar to that of p255/2, but in which the nucleotide in position -3 was changed to A (p255/15) or G (p255/16). The mutagenesis protocol is described in detail elsewhere (7). p255/15 was subsequently used to generate mutant p255/15A by the same procedure whereby p255/2A was derived from p255/2 (see above). Similarly, p255/16 was used to generate mutant p255/16A.

Construction of Plasmids p255/20 and p255/21

p255 was digested with HindIII and BamHI, releasing a 178 bp fragment

(see Figure 1a) which includes the ATG triplet at the start of the preproinsulin coding sequence. The gel-purified fragment was incubated with DNA ligase and a 100-fold molar excess of the 5'-phosphorylated oligonucleotide 5'-GATC-(C,T)AAGCTT(G,A)-3', which was custom-synthesized by PL Biochemicals, Inc. The GATC sequence at the 5'-end of the oligonucleotide is complementary to the sequence at the Bam-cut end of the 178 bp fragment. The product of the ligase reaction was digested with HindIII, which cuts within the adaptor at the sequence A/AGCTT. The DNA was then fractionated by polyacrylamide gel electrophoresis, and a 184 bp fragment was recovered [i.e., the original 178 bp fragment modified at the right end by the addition of GATC(C,T)A]. The 184 bp fragment was inserted into the unique HindIII site of p255, generating p255/20 (in which the sequence at the right end of the insert was GATC*C*A) and p255/21 (in which the sequence at the right end of the insert was GATC*T*A).

Construction of Plasmid p255/9

The derivation of this mutant is a bit circuitous. The objective was to reiterate the ribosome binding site of the preproinsulin gene in a way that would place the upstream ATG triplet out-of-frame with respect to the preproinsulin coding sequence that lies downstream. In order to do this inexpensively, I used two derivatives of p255 that were already on hand. From the previously-described mutant p255/8 (ref. 3), a 185 bp fragment was obtained by digesting with HindIII. That fragment includes the ATG codon that normally initiates synthesis of preproinsulin. In the sequence shown for p255/9 in Figure 5, the 185 bp HindIII fragment extends from the left end of the insert up to the sequence AAGCTT within the boxed region at the right end of the insert. The remaining nucleotides in the boxed region were acquired by inserting the 185 bp fragment *not* into wild type p255, but into the HindIII site of mutant p255/2, which already carries an 11 bp insertion. The sequence of the entire 196 bp insertion in p255/9 is shown in Figure 5.

Derivation of Mutants C1 and C2

For these constructions I began with p255/11 (ref. 7), which is identical to wild type p255 except for absence of the BamHI site at the junction between rat genomic and pBR322 sequences. As illustrated in the upper portion of Figure 7, p255/11 was digested with HindIII and BamHI, eliminating a 178 bp fragment that includes the natural ATG initiator codon for preproinsulin. The synthetic oligonucleotide $\begin{matrix} 5'\text{-AGCTACC}\overset{A}{G}\text{TGGTACCATGGG-3'} \\ 3'\text{-TGGTACCATGGTACCCCTAG-5'} \end{matrix}$ was then inserted by making use of the HindIII (AGCT) and BamHI (GATC) overhangs at the left and right ends, respectively. In mutant C2, the inserted oligonucleotide has two ATG codons flanked by identical sequences (TACC*ATGG*); the 5'-proximal ATG is out-of-frame with respect to the preproinsulin coding sequence, while the second

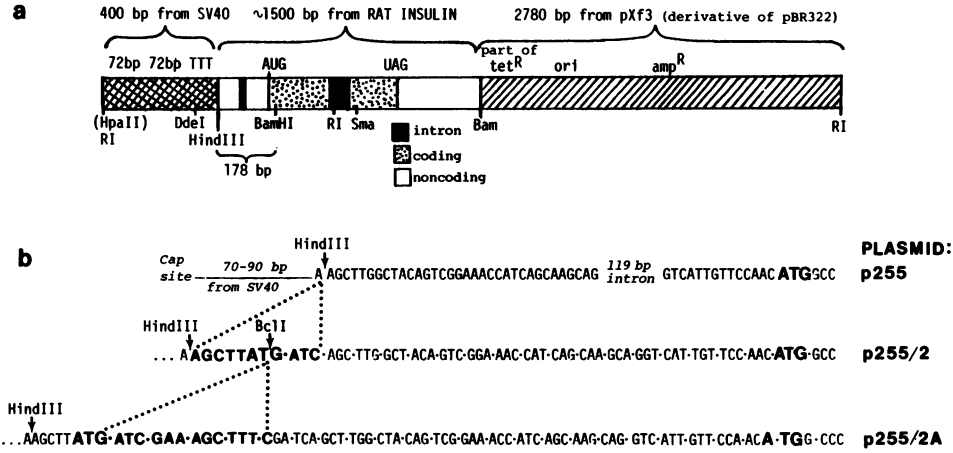


FIGURE 1. Structure of p255 and derivative plasmids.
(a) Linear representation of the parental plasmid p255, in which transcription of the rat preproinsulin gene is under the control of the SV40 early promoter. The plasmid contains a unique HindIII site at the boundary between SV40 and rat insulin DNA. Transcription initiates 70-90 bp upstream from that site.
(b) Construction of derivatives p255/2 and p255/2A. Sequences are shown for the portion of the DNA that corresponds to the 5'-noncoding region of mRNA. The ATG triplet shown in boldface in the parental plasmid p255 (top line) is the site where translation of preproinsulin normally initiates. p255/2 was obtained by inserting, at the HindIII site of p255, the 11 bp oligonucleotide AGCTTATGATC (shown in boldface in line 2; see Experimental Procedures). p255/2A was obtained by inserting, at the BclI site of p255/2, the 14 bp oligonucleotide GATCGAAAGCTTTC, which is shown in boldface in line 3. The sequences of p255/2 and p255/2A are punctuated in the reading frame set by the 5'-proximal ATG triplet, and the sequences are shown without the 119 bp intron that interrupts the 5'-noncoding region.

ATG triplet is in-frame. The control plasmid C1 has a single ATG triplet which is in-frame and functions as an initiator codon for preproinsulin (see Fig. 7).

## RESULTS

### Insertion and Mutagenesis of a Synthetic ATG-Containing Oligonucleotide Upstream from the Preproinsulin Coding Sequence

The starting point for these constructions was a shuttle vector called p255 which carries the rat preproinsulin II gene linked to the SV40 early promoter. That plasmid, which was generously given to me by Dr. Peter Lomedico, has been described previously (3,12); it is shown schematically in Figure 1a. There are no AUG triplets upstream from the preproinsulin start site in the transcript derived from wild type p255. To determine whether an AUG codon introduced upstream from the preproinsulin coding sequence would serve as a

"barrier," preventing ribosomes from initiating at the natural start site, I inserted into p255 an 11 bp synthetic oligonucleotide that carries an ATG triplet. The structure of the resulting mutant (p255/2) is shown in Figure 1b, and the details of the construction are given in Experimental Procedures. Because the upstream AUG triplet in p255/2 lies in the same reading frame as the preproinsulin coding sequence, I expected to be able to monitor initiation at the first versus the second AUG codon by measuring the size of the insulin-related polypeptides: the polypeptide resulting from initiation at the upstream AUG would have 18 extra amino acids at the N-terminus. Preliminary experiments revealed, however, that the "long form" of preproinsulin encoded by constructs analogous to p255/2 underwent cleavage by signal peptidase. Thus, irrespective of which AUG codon was used for initiation, the only polypeptide that accumulated was normal-sized proinsulin. A more workable approach for assessing the degree to which an upstream AUG triplet can supplant initiation at the normal start site was to place the upstream AUG triplet in a different reading frame from the preproinsulin coding sequence, as shown in Figure 1b for derivative p255/2A. Although initiation at the upstream AUG triplet cannot be measured directly in such constructs, the extent to which the upstream AUG serves as a barrier to initiating downstream can be assessed by monitoring the yield of proinsulin. This is the underlying design in all of the experiments that follow.

For reasons alluded to in the Introduction, it was of interest to vary the nucleotide sequence around the upstream ATG triplet in p255/2A. Targeted mutagenesis of the sequence preceding the initiator codon was accomplished by annealing a mismatched oligonucleotide to the single-stranded region in a gapped heteroduplex circle. The general mutagenesis procedure and the explicit protocol for generating derivatives p255/15 and p255/16 have already been described (7). p255/2 (diagrammed in Figure 1b), p255/15 and p255/16 differ from one another in one position only: whereas cytidine occurs 3 nucleotides upstream from the inserted ATG triplet in p255/2, p255/15 has adenosine and p255/16 has guanosine in that position. Derivatives p255/15A and p255/16A--in which the upstream ATG triplet has been shifted out-of-frame with respect to the preproinsulin coding sequence--were obtained from p255/15 and p255/16, respectively, by a procedure identical to that used for the derivation of p255/2A from p255/2 (see Figure 1b and Experimental Procedures). The structures of p255/2A, 15A and 16A were confirmed by DNA sequence analysis. The relevant portions of the sequencing gels are shown in Figure 2.

In summary, derivatives p255/2A, 15A and 16A have a 25 bp insertion at the HindIII site of p255. The AUG triplet carried within the insert is out-

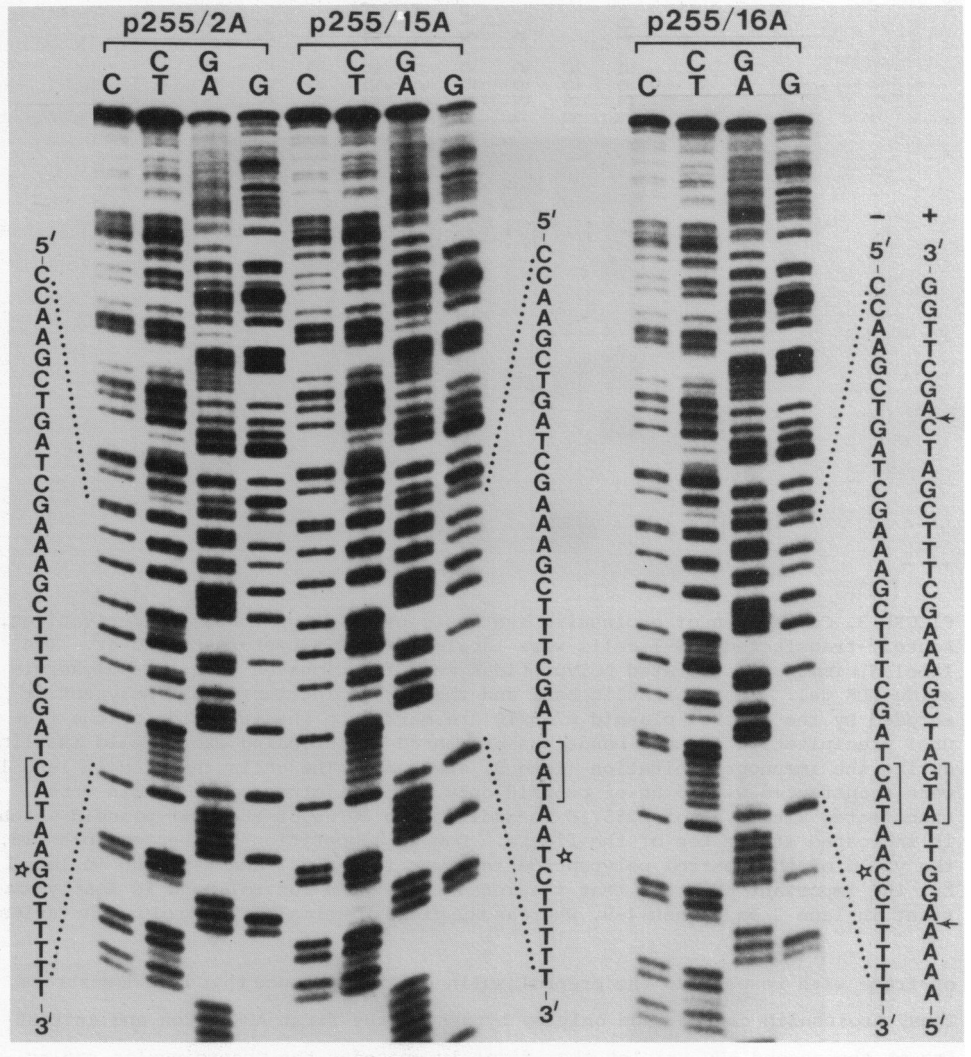


FIGURE 2. DNA sequence analysis of p255/2A, p255/15A and p255/16A. The anti-sense (minus) strand of the 256 bp DdeI-BamHI fragment from each plasmid was labeled at the 3'-end with $^{32}P$ and subjected to sequence analysis by the chemical cleavage method (9). The DdeI site that carries the $^{32}P$ label lies in the SV40-derived region of the plasmid, 52-56 bp upstream from the unique Hind III site. For convenience, the sequence of the sense (plus) strand is also written out for p255/16A. Arrows mark the boundaries of the 25 bp insert in p255/16A. In each sequence the ATG triplet or its complement is bracketed. The single nucleotide difference between p255/2A, p255/15A and p255/16A is starred.

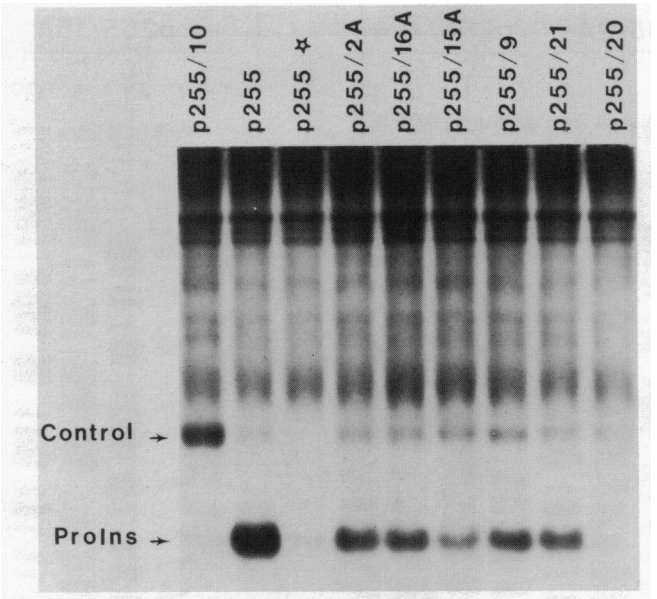


FIGURE 3. Comparison of proinsulin synthesis by wild type p255 and derivatives. Acutely-transfected COS-1 cells were labeled with $^{35}$S-cysteine for 3 hr. The labeled, immunoprecipitated polypeptides were fractionated on a 15% polyacrylamide-SDS gel. The proinsulin band and the high molecular weight polypeptide encoded by the control plasmid p255/10 are marked on the left side of the figure; precipitation of both bands was prevented by including excess cold insulin during the immunoprecipitation (lane 3, starred). The cells analyzed in lane 1 were transfected with 5 µg of p255/10 only. In all other cases, cells were transfected with 5 µg of p255/10 in addition to 20 µg of the test plasmid which is indicated at the top of the figure. Due to competition by the test plasmid, the yield of the control polypeptide is lower in lanes 2 and 4-9 than in lane 1; but the important point is that the intensity of the control band is nearly constant in lane 2 and lanes 4-9, whereas the yield of normal-sized proinsulin varies.

of-frame with respect to the preproinsulin coding sequence that lies downstream. Thus, proinsulin can be made only by bypassing the first AUG codon and initiating at the second AUG triplet that directly precedes the preproinsulin coding sequence. The extent to which the upstream AUG triplet serves as a barrier to the migrating 40S subunit can be assessed by the extent to which proinsulin synthesis is reduced in these constructs, relative to wild type p255. The single nucleotide difference between p255/2A, 15A and 16A was introduced to determine (in a preliminary way) whether the effectiveness of the barrier depends on sequences flanking the AUG codon.

Synthesis of Proinsulin from Modified Plasmids that Have an Upstream, Out-of-Frame ATG Triplet

A transient expression assay was used to compare the efficiency of trans-

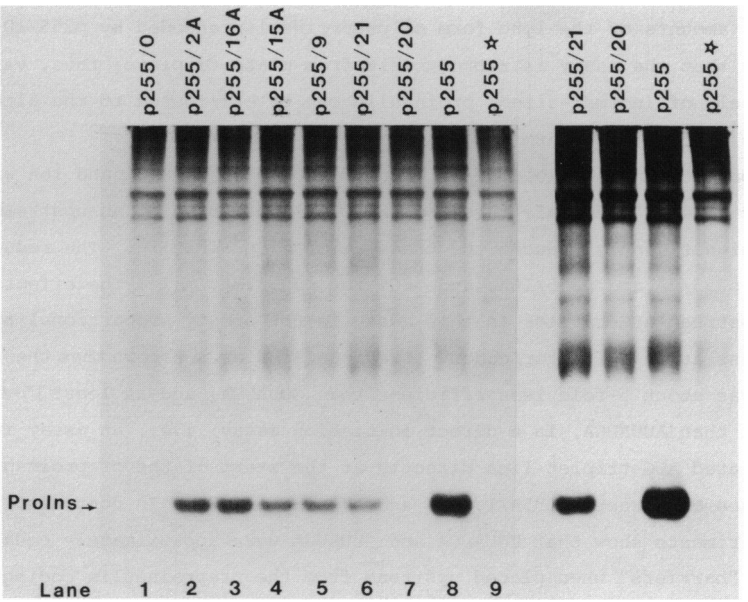


FIGURE 4. Variations in proinsulin synthesis resulting from inserting an AUG triplet upstream from the preproinsulin coding sequence.
COS-1 cells were transfected with 8 μg of one of the plasmids indicated at the top of the figure. Cells were labeled with $^{35}$S-cysteine for 4 hr. Plasmid p255/0, which lacks the entire preproinsulin coding sequence, was included as a negative control (lane 1). Inclusion of excess unlabeled insulin prevented the precipitation of the proinsulin band (lane 9, starred). At the right of the figure, lanes 6 to 9 are shown again after a longer exposure. For this experiment, the control plasmid p255/10 was omitted in order to increase the sensitivity of detecting very low levels of proinsulin. (Although p255/10 directs synthesis almost exclusively of the "long form," a small amount of the long form undergoes cleavage, generating a trace of proinsulin that would have contributed unwanted background.)

lating proinsulin from wild type p255 versus the insertion derivatives, p255/2A, 15A and 16A. Figures 3 and 4 show polyacrylamide gel analyses of $^{35}$S-labeled polypeptides extracted from COS cells 48 hr after transfection by the various plasmids. The transfection assay was carried out under conditions of plasmid DNA excess (Figure 3) as well as under conditions where the yield of proinsulin was proportional to the concentration of plasmid DNA (Figure 4). For the experiment shown in Figure 3, the cells were transfected with a mixture of the test plasmid (p255/2A, etc.) and a previously-described derivative called p255/10 (ref 3). The latter plasmid directs synthesis of a "long form" of preproinsulin which is readily resolved from authentic proinsulin, and hence could

be used as an internal control. The presence in lane 2 and lanes 4-9 of nearly identical amounts of the long form of preproinsulin encoded by p255/10 provides assurance that the assay is reproducible from plate to plate; thus, variations in the yield of (normal-sized) proinsulin can be attributed to the altered structure of the test plasmids.

Comparison of the yield of proinsulin from p255/2A, 15A and 16A with that from wild type p255 reveals three points: (i) Insertion of an upstream AUG triplet significantly reduces the translation of proinsulin. The reduction is ∿60% for p255/2A and p255/16A, and ∿80% for p255/15A. (ii) The effectiveness of the upstream-AUG-barrier in this assay is *not* simply proportional to its effectiveness as an initiator codon. Previous studies had shown that the sequence CUUAUGA was about 5-fold less efficient than GUUAUGA, and at least 15-fold less efficient than AUUAUGA, in a direct initiation assay; i.e., an assay in which the indicated AUG triplet lies directly at the start of the preproinsulin sequence (see constructs p255A/1, A/4 and A/3 in ref. 7). In contrast, the present experiments show that CUUAUGA and GUUAUGA were approximately equally efficient as "barriers" when placed upstream from the preproinsulin coding sequence (Figure 3, lanes 4 and 5) and AUUAUGA was only slightly more effective as an upstream barrier (Figure 3, lane 6). This issue will be discussed more fully below. (iii) Although translation of proinsulin was significantly reduced in cells transfected by p255/2A, 15A and 16A, in no case was it abolished. This could be rationalized by postulating that the sequence around the upstream AUG triplet in those plasmids was not optimal for initiation--and, hence, not strong enough to stop the migration of all 40S subunits. (Note that, although p255/15A has the canonical A in position -3, it differs in every other position from the consensus sequence for eucaryotic initiation sites.) An alternative explanation, advocated by Lomedico and McAndrew (12), is that some ribosomes can bind directly to the preproinsulin start site without traversing the upstream region. In that case residual translation of proinsulin would persist no matter what sort of barrier were placed upstream. Direct binding of ribosomes at the preproinsulin start site seems to be ruled out, however, by the results described below with derivatives p255/20 and C2.

Construction and Expression of Plasmids in Which the Initiator Codon for Preproinsulin Is Repeated Upstream, Out-of-Frame with Respect to the Rest of the Preproinsulin Coding Sequence

As explained in the preceding section, a reasonable hypothesis is that the upstream AUG triplet in p255/2A, 15A and 16A prematurely halts the migration of some 40S subunits, thereby reducing the number of 40S subunits that reach the preproinsulin

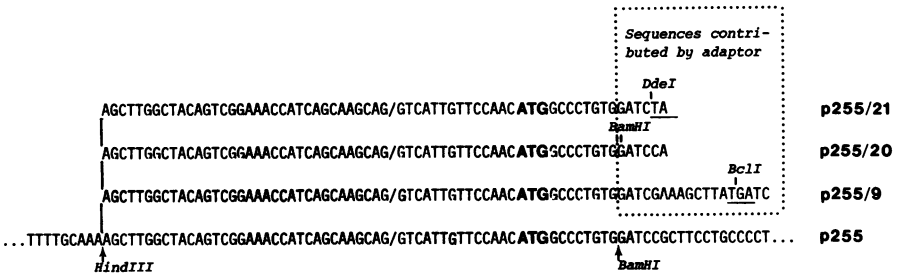


FIGURE 5. Structure of plasmids in which the initiator codon for preproinsulin is repeated upstream, out-of-frame with respect to the insulin coding sequence. The derivatives were obtained by inserting at the HindIII site of p255 a fragment of 196 bp (p255/9) or 184 bp (p255/20 and p255/21). Details of the constructions are given in Experimental Procedures. The sequences of each of the inserts, with the 119 bp intron represented by a diagonal line, are shown in the upper three lines of the figure. In the case of p255/21 and p255/9, ribosomes initiating at the ATG triplet carried within the insert would encounter a terminator codon (underlined) after translating just five (p255/21) or eight (p255/9) codons--well upstream from the start of the preproinsulin coding sequence. However ribosomes initiating at the ATG triplet carried within the insert in p255/20 would not encounter a terminator codon until far beyond the start of the preproinsulin coding sequence. The ATG triplet shown in boldface in the bottom line marks the beginning of the coding sequence for preproinsulin.

start site; but that the sequence context around the upstream AUG triplet is not good enough for the barrier-effect to be absolute. That explanation would be tenable if one could show that an initiator codon in a *highly favorable* sequence context would, when introduced upstream and out-of-frame, *completely* abolish translation of proinsulin. Plasmid p255/20 was constructed for this purpose. As shown in Figure 5, p255/20 was generated by duplicating a small fragment of DNA from the parental plasmid p255. The reiterated HindIII-BamHI fragment includes the initiator codon for preproinsulin in its natural context. The adaptor that was used to modify the right end of the insert in p255/20 was designed so that the ATG triplet carried within the insert would be out-of-frame with respect to the preproinsulin coding sequence that lies downstream. In that sense, p255/20 is similar in design to p255/2A, 15A and 16A: each has a single ATG triplet inserted upstream and out-of-frame with respect to the preproinsulin coding sequence. The result obtained when COS cells were transfected with p255/20 is shown in Figure 3, lane 9, and in Figure 4, lane 7: no proinsulin was made.

A different, and rather surprising, result was obtained with two other derivatives--p255/9 and p255/21. The structure of those plasmids is very similar to that of p255/20 *except for the short sequence contributed by the adaptor*

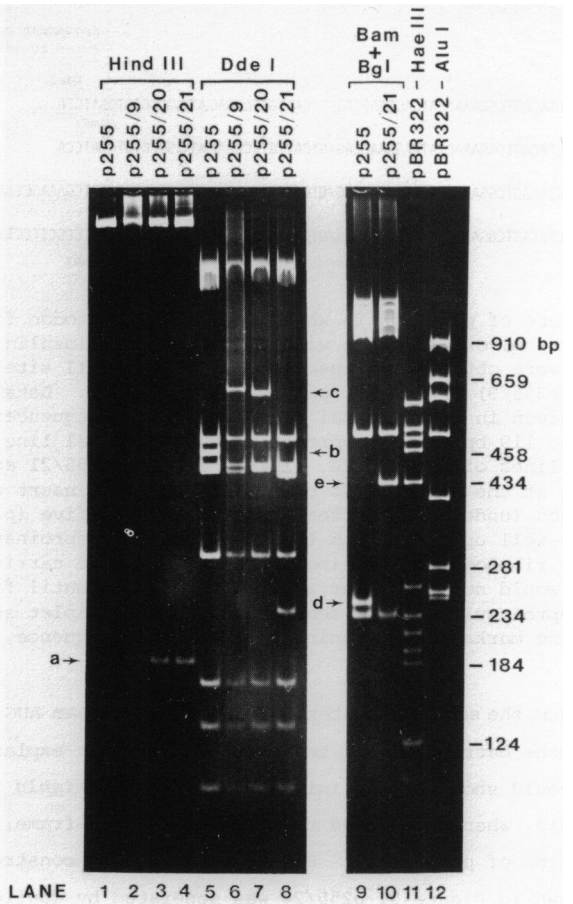


FIGURE 6. Restriction digestion patterns of p255 and the insertion derivatives p255/9, p255/20 and p255/21.
Ethidium bromide-stained DNA fragments were fractionated by electrophoresis in a 5% polyacrylamide gel, which was calibrated by reference to HaeIII and AluI fragments of pBR322 (lanes 11, 12). The sizes of some marker bands are shown on the right side of the figure. Digestion of p255/20 and p255/21 with HindIII yields a fragment of 184 bp (band a, lanes 3 and 4), consistent with the sequences shown in Figure 5. A slightly larger fragment is released from p255/9 upon digestion with HindIII (lane 2). To establish that each plasmid carries a single copy of the 184 bp insert (or 196 bp in p255/9), the DNA's were digested with DdeI. The control pattern obtained with wild type p255 is shown in lane 5. Digestion of p255/9 or p255/20 with DdeI (lanes 6 and 7, respectively) yielded the same profile as p255 except for the absence of a 460 bp fragment, the position of which is marked by arrow b. Instead of the 460 bp band, the DdeI profiles of p255/9 and p255/20 have a fragment of ∿650 bp (arrow c). The size of the new band is approximately the sum of the 460 bp parental fragment plus the 184- or 196 bp insert. A corresponding band was not obtained with p255/21 (lane 8) because the adaptor used in the construction of that plasmid contributes a new DdeI cleav-

age site; thus the 650 bp band is replaced by two smaller fragments. As an alternative to using DdeI, p255/21 was digested with BamHI and BglI. Comparison of the patterns thus obtained with p255/21 and wild type p255 reveals a single difference: the ∿250 bp fragment in lane 9 (wild type, band d) has been replaced by a ∿434 bp fragment (lane 10, band e), consistent with the presence of a single copy of the 184 bp insert in p255/21.

(see Figure 5). p255/9 was constructed merely as an alternative to p255/20; i.e., another example in which the strong initiator codon of the preproinsulin gene was tested as an upstream "barrier." Unlike p255/20, however, p255/9 was found to direct synthesis of a modest level of proinsulin (Figure 3, lane 7; Figure 4, lane 5). Inspection of the sequence of p255/9 revealed a translation terminator codon (UGA) 8 codons beyond the 5'-proximal AUG triplet. This raised the possibility that ribosomes might initiate at the first (out-of-frame) AUG triplet and then terminate, releasing the 60S subunit but retaining the 40S subunit which could *reinitiate* at the start of the preproinsulin coding sequence. The last plasmid in this series, p255/21, provides strong evidence that presence or absence of a terminator codon indeed determines whether ribosomes can initiate at a downstream AUG triplet. p255/21 differs from p255/20 (described above) by only one nucleotide, as shown in Figure 5: in the case of p255/21 the sequence contributed by the adaptor ends in UA--which, when inserted at the HindIII site of p255, becomes the ochre terminator codon UAA. In the case of p255/20, however, the sequence contributed by the adaptor ends in CA, and thus no terminator codon occurs in-frame between the 5'-proximal AUG triplet carried on the insert and the AUG triplet at the start of the preproinsulin coding sequence. The structures of p255/21 and p255/20 were confirmed by digestion with appropriate restriction enzymes (Figure 6) and by DNA sequence analysis. When assayed for expression in COS-1 cells, p255/21 produced significant amounts of proinsulin (Figure 3, lane 8; Figure 4, lane 6) whereas p255/20 produced no proinsulin, as already noted.

Plasmids with an Out-of-Frame AUG Triplet Located Just Five Nucleotides Upstream from the Start of the Preproinsulin Coding Sequence

The observation that initiation at the second AUG triplet in p255/21 is dependent on the presence of a terminator codon between the first and second AUGs might be explained in either of two ways: (i) The explanation I favor is that ribosomes reach the second AUG triplet only after initiating at the 5'-proximal AUG, incorporating a few amino acids and then terminating; the 40S ribosomal subunit remains bound to the mRNA during termination, subsequently advances, and reinitiates at the second AUG triplet. According to this mechanism (which is a variation of the scanning model) 40S subunits cannot bind di-

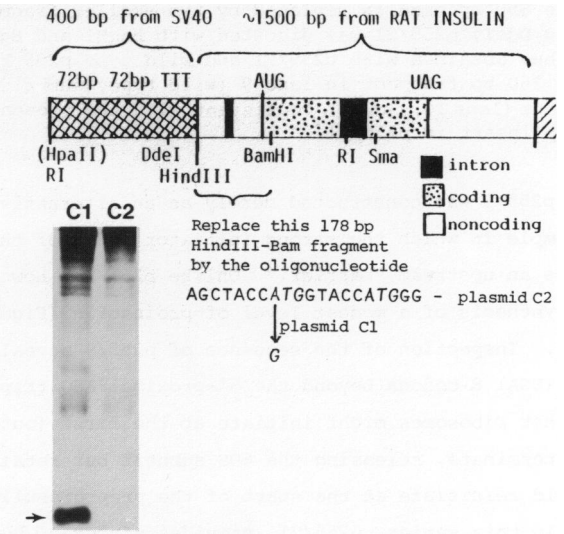


FIGURE 7. Structure and expression of mutants C1 and C2.
A portion of the parental plasmid p255/11 is shown in the top line. In C1 and C2, the indicated 20 bp oligonucleotide was inserted in place of the HindIII-BamHI fragment that carries the natural initiator codon for preproinsulin. In mutant C1, the oligonucleotide insert carries a single ATG codon, which is in-frame with the preproinsulin coding sequence and is a functional initiator codon. The polyacrylamide gel profile at the left documents that proinsulin (marked by an arrow) is synthesized in COS cells transfected by the control plasmid C1. In mutant C2, the oligonucleotide insert carries two ATG triplets: the first is out-of-frame and the second is in-frame with the preproinsulin coding sequence.

rectly to internal AUG codons. Moreover, the inability to initiate at the downstream preproinsulin start site in constructs like p255/20 (i.e., constructs which lack a terminator codon between AUG-1 and AUG-2) should persist no matter how long or short the distance between the two AUGs. (ii) An alternative explanation for the difference in expression between p255/20 and p255/21 is that 40S subunits *can* bind directly to internal AUG codons, but are prevented from doing so in p255/20 because the downstream initiation site is blocked by a stream of 80S ribosomes advancing from the upstream start site. The terminator codon in p255/21 would halt the 80S ribosomes, thereby facilitating direct, independent initiation (not "reinitiation") at the downstream AUG codon. According to this "80S blockade" model, the inhibitory effect of the upstream out-of-frame AUG codon should disappear (without the necessity of introducing a terminator codon) when the second AUG is placed very close to the first. In a construct such as C2 (Figure 7), in which an out-of-frame AUG codon lies just five nucleotides upstream from the AUG triplet that heads the preproinsulin coding sequence,

*there is simply no room for an "80S blockade" to mask the second AUG while leaving the first exposed.* Thus, mechanism ii predicts that mutant C2 *should* make some proinsulin. The scanning model (embodied in mechanism i above) predicts that C2, like p255/20, should make no proinsulin. The observed result is that C2 makes no proinsulin (Figure 7). The matched control plasmid C1, which has GUG rather than AUG in the upstream position, directs synthesis of a normal level of proinsulin (Figure 7).

## DISCUSSION

### The Presence of an Upstream AUG Triplet Reduces Initiation at the Downstream Start Site for Preproinsulin

The experiments described above confirm the prediction that upstream AUG triplets function as barriers, diverting ribosomes from potential initiation sites that lie farther downstream. Thus, the yield of proinsulin was significantly reduced with all of the mutants that I tested. The insertion mutants previously analyzed by Lomedico and McAndrew (12) were similarly impaired in their ability to synthesize proinsulin. For two reasons, however, the barrier-effect of upstream AUG triplets was not absolute with many of the constructs used in this study. (i) The effectiveness of an upstream AUG triplet (i.e., the tendency of 40S ribosomal subunits to stop there, rather than advance) depends on the sequence context around the AUG codon. This is shown by the difference in residual proinsulin synthesis between p255/2A (∿40% of the wild type level), p255/15A (∿20% residual synthesis) and p255/20 (no residual synthesis). (ii) Even when an upstream AUG triplet occurs in an optimal context, such that virtually all 40S ribosomal subunits stop and initiate there, initiation at a downstream AUG codon is not precluded if the ribosomes encounter a terminator codon before they reach the second potential initiation site. The simplest interpretation is that eucaryotic ribosomes can reinitiate following a terminator codon. Thus, in a message that has an appropriate arrangement of initiator-terminator-initiator codons, as in p255/21, ribosomes can reach downstream coding sequences that would be otherwise inaccessible. A few anecdotal reports of reinitiation in eucaryotes have appeared during the past year or so (13, 14), but the analysis of proinsulin synthesis by p255/20 and p255/21 seems to be the most systematic demonstration: a single nucleotide difference between the two plasmids that abolishes the terminator codon abolishes the expression of proinsulin. That unexpected finding requires some further comment.

Evaluation of the Hypothesis that Eucaryotic Ribosomes Can Reinitiate

*Reinitiation versus Alternative Explanations*

An alternative explanation for the residual synthesis of proinsulin by p255/21 is that ribosomes can bind (albeit inefficiently) *directly* to the AUG triplet that precedes the preproinsulin coding sequence. One could argue that the downstream initiation site is inaccessible in the control plasmid p255/20 due to masking of that region by a stream of 80S ribosomes advancing from the strong upstream initiation site, or due to conformational constraints; in p255/21 the constraints would be relieved when 80S ribosomes pause at the upstream terminator codon. But various considerations argue against such ideas: (i) The inability of eucaryotic ribosomes to bind to circular templates (15, 16) contradicts the "direct binding" hypothesis. The translational properties of preproinsulin-encoding mRNAs that contain tandemly-reiterated ribosome binding sites (ref. 3) also seem to rule out a mechanism in which ribosomes bind and initiate directly at internal AUG codons. (ii) There is no evidence that conformational constraints are responsible for the usual inability of eucaryotic ribosomes to initiate at internal AUG codons. On the contrary, efforts to activate internal initiation sites by denaturing plant and animal viral RNAs have given uniformly negative results (17-19). Although it is easy to show that mRNA conformation regulates the access of *procaryotic* ribosomes to potential initiation sites, there is no experimental justification for extrapolating that hypothesis to eucaryotes. It follows that, although the "reinitiation" phenomenon in procaryotes might be explained by postulating that pausing of a ribosome at a terminator codon perturbs the structure of the mRNA in a way that gives other ribosomes access to potential initiation sites that lie downstream (20, 21), it is far less reasonable to invoke such a mechanism for reinitiation in eucaryotes. (iii) The notion that a stream of 80S ribosomes advancing from an upstream initiation site might interfere to some extent with initiation at downstream sites could not be dismissed a priori at the outset of this study, although it seemed unlikely that such a mechanism would *completely* suppress downstream initiation--as was observed with p255/20. [In procaryotic systems, the advance of 70S ribosomes from an upstream initiation site often reduces, but never abolishes, the translation of an overlapping gene (22-24).] However, the experiment illustrated in Figure 7 with mutant C2 discredits the idea that an 80S blockade is responsible for suppressing initiation at downstream AUG codons. In plasmid C2, the out-of-frame AUG codon lies only five nucleotide upstream from the second (in-frame) AUG triplet; because of their proximity, it seems far-fetched to think that 80S ribosomes could completely block access to

the second AUG but not the first. But the second AUG triplet is *not* functional in mutant C2; no proinsulin is made. That very clear result obtained with mutant C2 constitutes strong evidence for "scanning," against "direct internal initiation," and against the idea that an "80S blockade" accounts for the failure of p255/20 to make proinsulin. (Were that the explanation for p255/20, one would have to find yet another mechanism to explain the failure of C2 to make proinsulin.)

*The Conclusion that Eucaryotic Ribosomes Can Reinitiate Is Not Necessarily at Odds with Older Data*

The emerging hypothesis that eucaryotic ribosomes can reinitiate following a terminator codon seems, at first glance, to be contradicted by a considerable body of older data obtained with plant and animal viral mRNAs. Although the literature is replete with experiments that seemed to prove the inability of eucaryotic ribosomes to translate 3'-proximal cistrons in polycistronic viral mRNAs (reviewed in ref. 25), nearly all of those experiments were carried out in vitro; it is not unreasonable to think that dilute cell-free extracts do not adequately reproduce the conditions that permit reinitiation to occur in vivo. Even in vivo, the efficiency of reinitiating at the 3'-proximal cistron in polycistronic viral mRNAs might be low (as it was in the present experiments with p255/9 and p255/21), thus hindering its detection. It would obviously be a mistake to cite the mere existence of subgenomic mRNAs as evidence that eucaryotic ribosomes cannot express downstream cistrons: although reinitiation might allow low-level translation of 3'-proximal cistrons in some polycistronic viral mRNAs (see below for which ones), *efficient* translation would still require a subgenomic mRNA in which the cistron in question was nearer the 5'-end.

With some polycistronic viral messages, the possibility of reinitiation is precluded by the structure of the mRNA: the beginning of the 3'-proximal cistron overlaps the end of the 5'-proximal cistron in the case of polyoma virus 19S mRNA (in which the coding sequence for VP1 is present but not expressed--ref. 26) and the genomic mRNAs of several retroviruses (in which the envelope protein sequence is present but not translated--refs. 27-29). In a few other cases, the 5'- and 3'-proximal cistrons do not overlap; but the presence of several AUG triplets within the intercistronic region might attenuate the movement of 40S ribosomal subunits through that region (30-32). With certain other polycistronic viral mRNAs (33-35), however, the sequence of the intercistronic region appears as if it would be compatible with reinitiation at the start of the 3'-proximal cistron. Perhaps reinitiation is the explanation behind the recent finding (36) that in vivo synthesis of the coat protein of al-

falfa mosaic virus begins prior to detectable synthesis of RNA-4, which was generally thought (from in vitro studies) to be the only functional message for coat protein.

Some of the yeast cytochrome c mutants whose translation in vivo has been studied by Sherman and Stewart (37) seemed to rule out the possibility that eucaryotic ribosomes reinitiate following a terminator codon. Their findings might reflect a difference between yeast and higher eucaryotes. Alternatively, reinitiation might require certain features--such as an optimal distance between the terminator and the next potential initiator codon--that were absent from the yeast mutants but fortuitously present in the insulin plasmids that I worked with.

*Reinitiation Solves Some Old Puzzles and Raises Some New Possibilities*

The ability of eucaryotic ribosomes to reinitiate following a terminator codon provides a tenable mechanism for the expression of a few problematical viral messages--such as poliovirus genomic RNA (38 and references therein) and the genomic (27) and subgenomic mRNAs of Rous sarcoma virus (RSV) (39). In each of those messages the 5'-"untranslated" sequence includes one AUG triplet in a highly favorable context for initiation. The presence of additional "weak" AUGs near the 5'-end of those mRNAs poses no special problem, but few if any 40S subunits would be expected to slip past the strong upstream AUG triplet. In each case, however, ribosomes that initiate at that AUG codon would terminate before reaching the AUG triplet that heads the major open reading frame. Thus, I believe that ribosomes gain access to the major downstream coding sequence in RSV and poliovirus mRNAs by initiating, terminating, and then reinitiating. A similar explanation might underlie the ability of SV40 late 16S mRNA to direct synthesis of both the agnoprotein (encoded near the 5'-end) and the major capsid protein VP1 which is encoded downstream (40).

The phenomenon of reinitiation might have broader implications than merely providing a solution for the expression of a few unusual viral mRNAs. If 40S subunits remain bound at terminator codons and subsequently reinitiate during their transit across the 5'-end of a message, which is the explanation I favor for the experiments described herein, should 40S subunits not also remain bound after the 80S ribosome has translated the major protein coding sequence and arrived at the terminator codon near the 3'-end of the message? The subsequent migration of 40S subunits through the 3'-"noncoding" region could have various interesting consequences, one of which is the possibility of recycling 40S ribosomal subunits back to the 5'-end of the message that has just been translated.

By What Mechanism(s) Do "Weak" Upstream AUG Triplets Reduce Translation of the Downstream Preproinsulin Sequence?

I attempted to answer this question by comparing the function of an AUG codon in a particular context (such as CUUAUGA) when the AUG triplet occurs (i) directly at the start of the preproinsulin coding sequence, with no other AUG triplets upstream (p255A/1 in ref. 7); (ii) upstream from the preproinsulin start site, and out-of-frame with respect to the preproinsulin coding sequence (p255/2A in this manuscript); and (iii) upstream from and in the same reading frame as the preproinsulin coding sequence (p255/2 in Figure 1). p255/2A and the related plasmids, p255/15A and 16A, reveal that the presence of an upstream out-of-frame AUG triplet interferes with initiation at the downstream preproinsulin coding sequence. Were the interference due solely to the upstream AUG triplet trapping some 40S ribosomal subunits, thereby reducing the number of 40S subunits that reach the preproinsulin start site, one would expect the extent of the interference to be proportional to the strength of that AUG triplet in the direct initiation assay: CUUAUGA--which, when positioned directly in front of the preproinsulin coding sequence in p255A/1, directs synthesis of a barely-detectable level of proinsulin--should, when placed upstream in a construct such as p255/2A, barely interfere with synthesis of proinsulin. That simple correlation was not seen in my experiments. The upstream, out-of-frame AUG triplet in p255/2A reduced the yield of proinsulin to the same extent as observed with p255/16A--although the upstream AUG triplet in the latter construct is in a much more favorable context, judging from the direct initiation assay (7).

One mechanism by which the interfering effect of the "weak" upstream AUG triplet in p255/2A could be amplified is if some 40S subunits detach from the message at such sites. That seems to be ruled out, however, by the finding that the yield of proinsulin is *not* reduced when an AUG triplet--again in the context CUUAUGA--occurs upstream and in the *same* reading frame as the preproinsulin coding sequence (unpublished data with p255/2). What mechanism might account for the inhibitory effect of a "weak" upstream AUG codon being amplified (relative to the efficiency with which it serves as an initiator codon) only when the upstream AUG triplet lies in a different reading frame from the preproinsulin coding sequence, as in p255/2A? Since ribosomes that initiate at the first AUG codon in p255/2A would continue to translate in the wrong reading frame for the full length of the preproinsulin mRNA, it is possible that an elongation barrier (such as the presence of a rare codon) could impede the progress of wrongly-phased ribosomes and, secondarily, of correctly-phased ribosomes attempting to synthesize preproinsulin from the same message. That possibility is under investigation.

Acknowledgments

I thank Peter Lomedico for providing the parental plasmid p255, Jim Pipas for cell culture facilities, and Harold Varmus for his comments on a preliminary version of the manuscript. This research was supported by a grant from the National Institutes of Health. M.K. is the recipient of Research Career Development Award AI00380.

REFERENCES

1. Kozak, M. (1980) Cell 22, 7-8.
2. Kozak, M. (1981) Current Topics in Microbiol. and Immunol. 93, 81-123.
3. Kozak, M. (1983) Cell 34, 971-978.
4. Kozak, M. (1981) Nucleic Acids Res. 9, 5233-5252.
5. Kozak, M. (1983) Microbiol. Reviews 47, 1-45.
6. Kozak, M. (1984) Nucleic Acids Res. 12, 857-872.
7. Kozak, M. (1984) Nature 308, 241-246.
8. Morrison, D.A. (1979) Methods in Enzymology 68, 326-331.
9. Maxam, A.M., and Gilbert, W. (1980) Methods in Enzymology 65, 499-560.
10. Gluzman, Y. (1981) Cell 23, 175-182.
11. Murphy, D., Brickell, P.M., Latchman, D.S., Willison, K., and Rigby, P.W.J. (1983) Cell 35, 865-871.
12. Lomedico, P.T., and McAndrew, S.J. (1982) Nature 299, 221-226.
13. Southern, P.J., and Berg, P. (1982) J. Mol. Appl. Genet. 1, 327-341.
14. Crowley, C., Liu, C., and Levinson, A. (1983) Mol. Cell. Biol. 3, 44-55.
15. Kozak, M. (1979) Nature 280, 82-85.
16. Konarska, M., Filipowicz, W., Domdey, H., and Gross, H.J. (1981) Eur. J. Biochem. 114, 221-227.
17. Collins, P.L., Fuller, F.J., Marcus, P.I., Hightower, L.E., and Ball, L.A. (1982) Virology 118, 363-379.
18. Monckton, R.P., and Westaway, E.G. (1982) J. Gen. Virol. 63, 227-232.
19. Zagorska, L., Chroboczek, J., Klita, S., and Szafranski, P. (1982) Eur. J. Biochem. 122, 265-269.
20. Kastelein, R.A., Berkhout, B., and van Duin, J. (1983) Nature 305, 741-743.
21. Steege, D.A. (1977) Proc. Natl. Acad. Sci. USA 74, 4163-4167.
22. Hoess, R.H., Foeller, C., Bidwell, K., and Landy, A. (1980) Proc. Natl. Acad. Sci. USA 77, 2482-2486.
23. Yen, T.S.B., and Webster, R.E. (1981) J. Biol. Chem. 256, 11259-11265.
24. Pollock, T.J., Tessman, E.S. and Tessman, I. (1978) J. Virol. 28, 408-410.
25. Smith, A.E. (1977) Federation of European Biol. Soc. Symp. 43, 37-46.
26. Smith, A.E., Kamen, R., Mangel, W.F., Shure, H., and Wheeler, T. (1976) Cell 9, 481-487.
27. Schwartz, D.E., Tizard, R., and Gilbert, W. (1983) Cell 32, 853-869.
28. Seiki, M., Hattori, S., Hirayama, Y., and Yoshida, M. (1983) Proc. Natl. Acad. Sci. USA 80, 3618-3622.
29. Shinnick, T.M., Lerner, R.A., and Sutcliffe, J.G. (1981) Nature 293, 543-548.
30. Ahlquist, P., Luckow, V., and Kaesberg, P. (1981) J. Mol. Biol. 153, 23-38.
31. Akusjärvi, G., and Persson, H. (1981) J. Virol. 38, 469-482.
32. Gould, A.R., and Symons, R.H. (1982) Eur. J. Biochem. 126, 217-226.
33. Barker, R.F., Jarvis, N.P., Thompson, D.V., Loesch-Fries, L.S., and Hall, T.C. (1983) Nucleic Acids Res. 11, 2881-2891.
34. Goelet, P., Lomonossoff, G., Butler, P., Akam, M., Gait, M., and Karn, J. (1982) Proc. Natl. Acad. Sci. USA 79, 5818-5822.

35. Ou, J., Rice, C., Dalgarno, L., Strauss, E., and Strauss, J. (1982) Proc. Natl. Acad. Sci. USA 79, 5235-5239.
36. Nassuth, A., Bruggencate, G., and Bol, J.F. (1983) Virology 125, 75-84.
37. Sherman, F., and Stewart, J.W. (1982) in Molecular Biology of the Yeast Saccharomyces: Metabolism and Gene Expression. Cold Spring Harbor Lab. Pp. 301-333.
38. Stanway, G., Cann, A., Hauptmann, R., Hughes, P., Clarke, L., Mountford, R., Minor, P., Schild, G., and Almond, J.W. (1983) Nucleic Acids Res. 11, 5629-5643.
39. Hackett, P.B., Swanstrom, R., Varmus, H.E., and Bishop, J.M. (1982) J. Virol. 41, 527-534.
40. Jay, G., Nomura, S., Anderson, C.W., and Khoury, G. (1981) Nature 291, 346-349.