MOLECULAR AND CELLULAR BIOLOGY, Oct. 1987, p. 3438–3445
0270-7306/87/103438-08$02.00/0

Vol. 7, No. 10

Effects of Intercistronic Length on the Efficiency of Reinitiation by Eucaryotic Ribosomes

MARILYN KOZAK

Department of Biological Sciences, University of Pittsburgh, Pittsburgh, Pennsylvania 15260

Received 6 April 1987/Accepted 3 July 1987

Simian virus 40-based plasmids that direct the synthesis of preproinsulin during short-term transfection of COS cells have been used to probe the mechanism of reinitiation by eucaryotic ribosomes. Earlier studies from several laboratories had established that the ability of ribosomes to reinitiate translation at an internal AUG codon depends on having a terminator codon in frame with the preceding AUG triplet and upstream from the intended restart site. In the present studies, the position of the upstream terminator codon relative to the preproinsulin restart site has been systematically varied. The efficiency of reinitiation progressively improved as the intercistronic sequence was lengthened. When the upstream "minicistron" terminated 79 nucleotides before the preproinsulin start site, the synthesis of proinsulin was as efficient as if there were no upstream AUG codons. A mechanism is postulated that might account for this result, which is somewhat surprising inasmuch as bacterial ribosomes reinitiate less efficiently as the intercistronic gap is widened.

Eucaryotic ribosomes usually initiate at the AUG codon that lies closest to the 5′ end of the mRNA (31). That tendency has been rationalized by postulating that the small ribosomal subunit binds initially at the capped 5′ end of the mRNA and subsequently migrates to the AUG initiator codon, which is recognized more or less efficiently depending on nearby sequences. From a comparison of several hundred vertebrate mRNAs, GCC$^{A}_{G}$CCAUGG has been proposed as the consensus sequence for initiation in higher eucaryotes (30, 33, Kozak, submitted for publication). The contribution of every nucleotide in that motif has been confirmed by mutagenesis (36; Kozak, J. Mol. Biol., in press). The most highly conserved nucleotides are the purine, most often A, in position −3 (i.e., three nucleotides upstream from the AUG codon) and the G in position +4; the aforementioned mutational analyses confirmed that those two positions have the strongest influence on initiation. Thus, a potential initiator codon can usually be designated as "strong" or "weak" by considering only those positions.

There are ways for eucaryotic ribosomes to initiate at internal sites in certain mRNAs, but those sites can be reached only by advancing from the 5′ end; never, it seems, by direct internal binding. An early indication of the constraint against direct internal binding was the inability of eucaryotic ribosomes to bind to circular RNA templates (27, 28), and a considerable body of other evidence has since been adduced (see Discussion). One mechanism by which ribosomes can reach an internal AUG codon is "leaky scanning," which occurs when the 5′-proximal AUG codon lies in an unfavorable context for initiation. Data consistent with leaky scanning have come from manipulating the sequences of synthetic constructs (35, 36, 40) and from analyzing naturally occurring viral mRNAs that are bifunctional (reviewed in references 38 and 39). A second mechanism for reaching an internal AUG codon operates when the upstream AUG codon (in a favorable context for initiation, thus ruling out leaky scanning) is followed shortly by an in-frame terminator codon. In such cases, ribosomes can initiate—or, more correctly, reinitiate—translation at the next AUG codon downstream (20, 35, 40, 51). Such mRNAs are formally bicistronic, although in most cases the upstream open reading frame is a "minicistron" that encodes only a small peptide. The efficiency of reinitiation in such constructs is usually low (35, 40).

This paper asks how reinitiation is affected upon varying the position of the upstream terminator codon relative to the start of the second open reading frame (ORF)—which, in this case, encodes preproinsulin. Whereas procaryotic ribosomes reinitiate most efficiently when the terminator codon that ends the first cistron overlaps the initiator codon of the second cistron (see Discussion), that is a very inefficient arrangement for reinitiation by eucaryotic ribosomes. Rather, with the chimeric preproinsulin mRNAs used for these studies, reinitiation progressively improved as the intercistronic distance increased. When the upstream minicistron terminated 79 or more nucleotides before the preproinsulin start site, ribosomes reinitiated with high efficiency, producing almost as much preproinsulin as if there were no upstream AUG codon. A mechanism is proposed that might explain this unexpected result.

MATERIALS AND METHODS

Construction of plasmids. The casette mutagenesis technique used for constructing the first-stage mutants (single-insert controls, Fig. 1B) has been described (36). Mutants were constructed by cleaving the acceptor plasmid (Fig. 1A) with *Hin*dIII and *Bam*HI and then inserting a partially double-stranded oligonucleotide that had at its left end a 4-nucleotide overhang complementary to the *Hin*dIII end of the acceptor and at its right end a 4-nucleotide overhang complementary to the *Bam*HI end of the acceptor. Only one strand of the oligonucleotide insert (the strand that has the same sequence as mRNA) is shown in Fig. 1B. The complementary oligonucleotides, purchased from Pharmacia P-L Biochemicals, Inc., were preannealed by incubating for 5 min at 85°C and then gradually cooling to room temperature. Subsequent ligation was at 14°C for 20 h. T4 DNA ligase and restriction enzymes were from New England BioLabs.

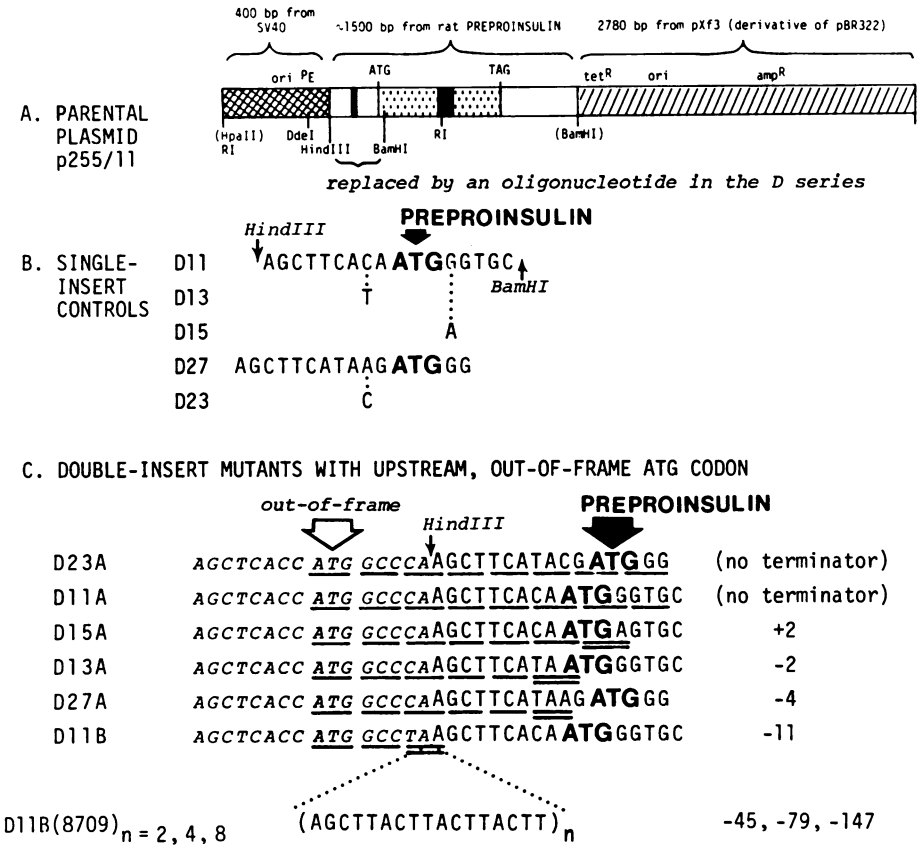

FIG. 1. Construction of mutants used to evaluate the efficiency of reinitiation as a function of intercistronic distance. (A) The parental plasmid p255/11 has been described (34, 41). Cleavage of this plasmid at the unique *Hin*dIII and *Bam*HI sites releases a small fragment that contains the initiator codon for preproinsulin. In its place, synthetic ATG-containing oligonucleotides were inserted to obtain mutants in the D series. (B) In these control constructs the inserted oligonucleotide contains a single ATG codon that is in frame with, and serves as the initiator codon for, preproinsulin. D13 and D15 are identical to D11 except in the indicated positions; D23 differs from D27 in a single position, as shown. (C) These mutants were derived by linearizing the single-insert control constructs at their unique *Hin*dIII site and inserting a second oligonucleotide, the sequence of which is shown in italics. The start site for preproinsulin remains the same as in the single-insert controls (▼), but an out-of-frame ATG triplet (▽) now precedes the initiator codon for preproinsulin. In D23A and D11A the reading frame established by the upstream ATG codon terminates 214 (D23A) or 357 (D11A) nucleotides beyond the preproinsulin start site. In the remaining mutants, the upstream minicistron terminates in the vicinity of the preproinsulin initiator codon, as shown. In D11B(8709)$_{\times 2}$, D11B(8709)$_{\times 4}$, and D11B(8709)$_{\times 8}$ the sequence AGCTTACTTACTTACTT is reiterated two, four, and eight times, respectively, at the *Hin*dIII site of D11B, thereby moving the mincistron (ATG · GCC · TAA) progressively farther upstream from the preproinsulin start site.

The double-insert mutants (Fig. 1C) were obtained by inserting, at the *Hin*dIII site of a single-insert plasmid from stage one, the 16-base-pair oligonucleotide AGCTCACCAT GGCCCA (designated A) or the related sequence AGCTCA CCATGGCCTA (designated B).

Transformation of *Escherichia coli* HB101 or MM294 and subsequent screening, sequencing, and purification of plasmids were as described previously (32).

Expression of plasmids in COS cells. COS-1 cells (16), grown in Eagle minimal essential medium (no. 330-1430; GIBCO Laboratories) with 10% fetal calf serum, were used between the 3rd and 12th passage. Subconfluent monolayers of cells were transfected by the calcium phosphate method (71) as described previously (36). The cells were labeled with [^{35}S]cysteine 2 days after transfection, and the extracted proteins were immunoprecipitated and separated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis as described previously (32, 36). The yield of proinsulin (the processed form of the protein that accumulates in COS cells) was quantified after exposing fluorograms at −70°C for 5 to 25 days. Densitometry was performed on a densitometer (model 620; Bio-Rad Laboratories). Cytoplasmic mRNA concentrations were routinely monitored by dot-blot hybridization (36).

RESULTS

Control plasmids lacking an upstream ATG codon. Mutants in the D series were derived from p255/11, a shuttle vector that expresses preproinsulin from the simian virus 40 early promoter (Fig. 1A). Construction of these mutants was carried out in two stages. The sequences of the first-stage, single-insert intermediates (Fig. 1B) were designed to fulfill two functions: the inserted oligonucleotide contains an ATG codon that serves as the new start site for synthesis of preproinsulin, and the insert contains a TAA or TGA codon that will halt translation coming from the upstream ATG codon in the second-stage derivatives. The single-insert controls themselves have no spurious ATG codons preceding the preproinsulin start site. (The initiator codon will be written as ATG for the remainder of this paper, inasmuch as all recombinant techniques and sequencing were performed

at the DNA level.) A preliminary transfection experiment revealed that the minor sequence differences among the controls (those differences determine precisely where upstream translation will terminate in the second-stage mutants) did not significantly affect the yield of proinsulin (Fig. 2). Because mutants D11, D13, D15, D23, and D27, in which reinitiation is not a factor, thus support translation with the same high efficiency, any differences in the yield of proinsulin among the double-insert mutants described next can be attributed to differences in the efficiency of reinitiation. No significant variation in mRNA levels was observed among mutants in the D series, and thus all of the differences in expression reflect translational events.

Effects of intercistronic length on the efficiency of reinitiation. Second-stage mutants D11A, D11B, D13A, D15A, D23A, and D27A were obtained by inserting another ATG-containing oligonucleotide upstream from the first (Fig. 1C). These derivatives consequently have a strong, out-of-frame ATG codon preceding the preproinsulin start site, and my expectation from previous studies (see above) was that synthesis of proinsulin would require a terminator codon to halt translation from the upstream site and allow reinitiation at the preproinsulin start site. The controls D23A and D11A, in which the reading frame set by the upstream ATG triplet continues far beyond the preproinsulin start site, indeed made no detectable proinsulin (Fig. 3). All of the other second-stage mutants were able to synthesize proinsulin, but the yield, which reflects the efficiency of reinitiation, varied at least 10-fold (Fig. 3). Reinitiation was remarkably inefficient when the upstream minicistron terminated close to or overlapped the ATG codon that initiates preproinsulin (D15A, D13A, D27A, and D11B in Fig. 3). In contrast, D11B(8709)$_{\times 8}$, in which the minicistron has been moved far upstream (Fig. 1C), synthesized proinsulin almost as efficiently as D11, which has no upstream ATG codons. Quantitation of appropriately exposed fluorograms similar to the one in Fig. 3 indicated efficiencies of ~70% for

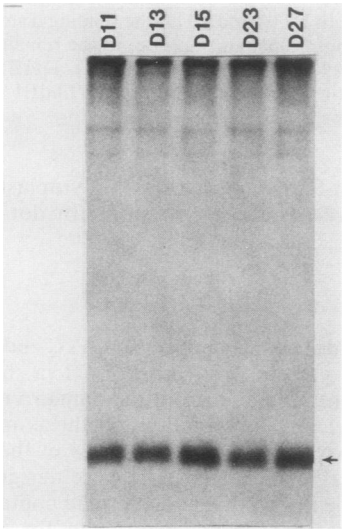

FIG. 2. Efficiency of translation does not vary significantly among the single-insert control plasmids. A fluorogram is shown of ^{35}S-labeled, immunoprecipitated proteins fractionated by polyacrylamide gel electrophoresis. The arrow indicates the position of proinsulin. Cells were pulse-labeled for 3 h immediately before harvesting. The lysate from 4×10^5 cells was analyzed in each lane. The sequence near the initiation site of each plasmid is given in Fig. 1B.

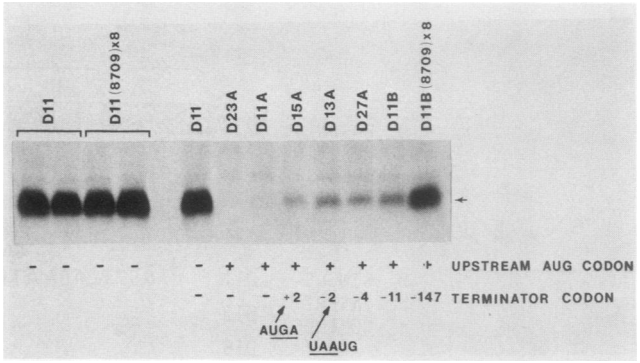

FIG. 3. Synthesis of proinsulin as a function of increasing the distance between an upstream minicistron and the preproinsulin start site. The control plasmids D11 and D11(8709)$_{\times 8}$ have no ATG codons upstream. D23A and D11A have an upstream, out-of-frame ATG codon that does not terminate before the preproinsulin start site. In the remaining mutants, the position of the terminator codon relative to the preproinsulin start site is as indicated (see also Fig. 1C). Adjacent lanes enclosed by brackets represent duplicate plates of cells that were transfected with the same plasmid. The fluorogram has been cropped; an arrow marks the position of proinsulin.

D11B(8709)$_{\times 8}$, 7 to 10% for D13A, D27A, and D11B, and ≤5% for D15A; the yield of proinsulin from D11 was taken as 100%.

As a control, eight copies of oligonucleotide 8709 were inserted into a plasmid that contains no upstream ATG codons and hence does not require that ribosomes reinitiate. In that case, lengthening the 5′-noncoding sequence by 136 nucleotides neither enhanced nor inhibited translation: D11(8709)$_{\times 8}$ directed synthesis of proinsulin with the same high efficiency as D11 (lanes 1 through 4, Fig. 3). Thus, the facilitating effect of the 136-nucleotide insertion in D11B(8709)$_{\times 8}$ occurs only when 40S ribosomal subunits are scanning in the reinitiation mode.

To determine more systematically how changes in intercistronic distance affect the efficiency of reinitiation, two, four, or eight copies of oligonucleotide 8709 were inserted at the *Hin*dIII site of mutant D11B. COS cells were transfected with this matched set of plasmids, and the yields of proinsulin were compared (Fig. 4). Inserting only two copies of the oligonucleotide, which increased the intercistronic length from 11 nucleotides (D11B) to 45 nucleotides, increased the efficiency of reinitiation from ~10% (D11B) to 20% [D11B(8709)$_{\times 2}$]. Inserting four copies of oligonucleotide 8709 lengthened the intercistronic gap to 79 nucleotides and allowed such efficient reinitiation that

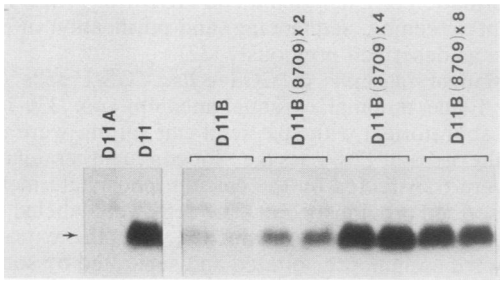

FIG. 4. Effect of systematically expanding intercistronic length. Details are given in the legend to Fig. 3. The positive control D11 has no upstream ATG codon; the negative control D11A has an upstream, out-of-frame ATG codon that does not encounter a terminator codon before the preproinsulin start site.

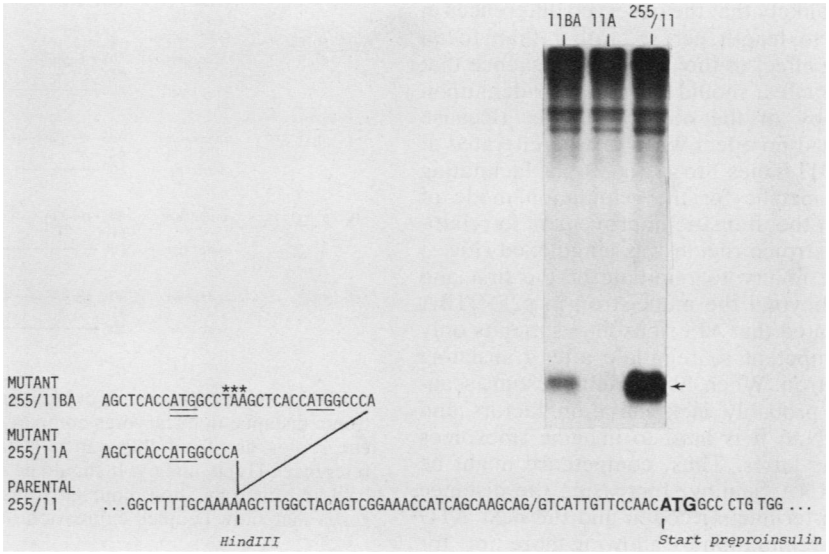

FIG. 5. Demonstration that reinitiation is not limited to the first ATG codon after a minicistron in the leader region of a chimeric mRNA that encodes preproinsulin. The complete structure of the parental plasmid p255/11 is given elsewhere (34). The portion of p255/11 reproduced here includes a *Hin*dIII site into which an oligonucleotide containing one or two ATG codons (as shown) was inserted, thereby generating p255/11BA or p255/11A, respectively. The insert shows a fluorogram of translation products synthesized in cells transfected by each plasmid. The proinsulin band is marked by an arrow. Whereas cells transfected by the control plasmid p255/11A made no proinsulin, the upstream minicistron in p255/11BA apparently relieved the inhibitory effect of the second, out-of-frame ATG codon, since preproinsulin synthesis (initiated now at the third ATG triplet) was partially restored.

D11B(8709)$_{\times 4}$ made as much proinsulin as the control plasmid D11. In the experiment shown, the yield of proinsulin declined slightly when the intercistronic distance was further lengthened to 147 nucleotides in D11B(8709)$_{\times 8}$, which might mean that the optimum lies somewhere between 79 and 147 nucleotides. In repeated experiments of this sort, however, the yield from D11B(8709)$_{\times 4}$ was sometimes less than shown in Fig. 4 (although never less than 50% of D11), and thus the difference in Fig. 4 between D11(8709)$_{\times 4}$ and D11(8709)$_{\times 8}$ might reflect experimental variation.

The inefficient reinitiation observed with D11B and most previously studied constructs (see above) might be explained in various ways. One possibility is that, when 80S ribosomes finish translating the upstream minicistron and 60S ribosomal subunits detach, some 40S subunits also detach; the efficiency of reinitiation would be determined by how many 40S subunits remain bound to the mRNA and resume scanning. However that mechanism cannot explain the dramatic improvement in reinitiation that occurs when the intercistronic sequence is lengthened. Another explanation for the usual inefficiency of reinitiation could be that, although 40S subunits remain bound after translating the upstream minicistron, some detach during the scanning phase before they reach the downstream ORF. In that case, reinitiation should become more efficient when the intercistronic gap is shortened, which is the opposite of what we have observed. The explanation I favor is that, after completing translation of the upstream minicistron, all 40S subunits resume scanning and they do not thereafter detach, but only some 40S subunits are competent to reinitiate when they reach the preproinsulin start site. Without defining at this time exactly what "competence" means, the general notion was tested in the following experiment.

Reinitiation is not limited to the closest ATG codon beyond the 5′-proximal ORF. If 40S ribosomal subunits only gradually become competent to reinitiate after translating the upstream minicistron, 40S subunits that are not yet competent when they reach the first ATG codon downstream should continue scanning, bypassing the first ATG and (as they eventually become competent) initiating at ATG codon(s) farther downstream. To test that prediction I constructed the plasmids shown in Fig. 5. The control plasmid p255/11 has no ATG codons upstream from the preproinsulin start site, and that plasmid accordingly synthesized the maximum amount of proinsulin. The negative control p255/11A has a strong upstream out-of-frame ATG codon that does not encounter a terminator codon before the preproinsulin start site; thus, the upstream ATG codon in p255/11A completely suppressed the synthesis of proinsulin, as expected. Mutant p255/11BA has an upstream ATG codon in the same context and same position as p255/11A (see the second ATG triplet in p255/11BA, which is underlined once), but that potentially inhibitory ATG is preceded by a minicistron, ATG · GCC · ***. Consequently, whereas 40S subunits are in the "primary" scanning mode when they encounter the upstream ATG codon in p255/11A, they are in the less efficient "reinitiation mode" when they encounter the same ATG codon in p255/11BA. In the latter case, not all 40S subunits were halted by the potentially inhibitory ATG codon; at least 15% reached the preproinsulin start site, judging from the yield of proinsulin shown in Fig. 5. If reinitiation were limited to the first ATG codon (in a good context) beyond the 5′-proximal ORF, the second ATG codon in p255/11BA should have been as effective a barrier as the first ATG codon in p255/11A, and neither plasmid should have been able to make proinsulin.

DISCUSSION

By reiterating a 17-base-pair oligonucleotide that was internally repetitious to begin with, the effects of intercistronic length could be studied without simultaneously introducing "new" sequences. The monotony of the expansion sequence in D11B(8709)$_{\times 2}$, D11B(8709)$_{\times 4}$, and

D11B(8709)$_{\times 8}$ makes it likely that the observed differences in translation were due to length per se, rather than to an inadvertent facilitating effect of the particular sequence that was inserted; the latter effect should have been evident upon inserting just one copy of the oligonucleotide. Because oligonucleotide 8709 had no effect when it was reiterated at the 5′ end of plasmid D11 (lanes 1 to 4, Fig. 3), the facilitating effect of length was specific for the reinitiation mode of translation. To explain the dramatic improvement in reinitiation when the intercistronic region was lengthened (Fig. 3 and 4) as well as the ability to reinitiate at the first and second ATG codons beyond the minicistron in p255/11BA (Fig. 5), I have postulated that 40S ribosomal subunits only gradually become competent to reinitiate after translating the upstream (mini)cistron. When 40S subunits resume scanning internally, they probably lack initiation factors and certainly lack Met-tRNA; it is hard to imagine ribosomes reinitiating without the latter. Thus, competence might be determined by Met-tRNA binding. Increasing the distance between the upstream terminator codon and the next ATG codon would improve reinitiation by allowing more time for Met-tRNA to bind. Conversely, shortening the intercistronic distance should cause Met-tRNA-deficient 40S subunits to bypass the nearest ATG codon and perhaps reinitiate farther downstream, as apparently happens with p255/11BA.

It seems unlikely that reinitiation is inefficient due to steric interference when the terminator and ATG codons overlap, as in D13A and D15A. The overlapping arrangement works beautifully in procaryotes (see below). Moreover, moving the terminator codon 45 nucleotides upstream from the preproinsulin start site, as in D11B(8709)$_{\times 2}$, should fully relieve any steric constraints; but D11B(8709)$_{\times 2}$ reinitiates only marginally better than D11B (Fig. 4) and not nearly as efficiently as D11B(8709)$_{\times 4}$.

An interesting possibility yet to be explored is that the efficiency of reinitiation might depend on the size of the upstream ORF. For the present study I used an upstream minicistron, partly because it seems easier to interpret results when the differences between constructs are kept small, and partly because long leader ORFs rarely occur in natural forms of eucaryotic mRNAs (38; however, see reference 69). The consequences of expanding the upstream ORF will have to be studied, however. A possible scenario is that some initiation factors remain transiently associated after the 60S subunit joins the 40S-mRNA complex at the first ATG codon, and the initiation factors are then released stochastically during elongation. If the transiently retained factors were needed for reinitiation, the ability to reinitiate should be inversely related to the size of the upstream ORF.

Recent experiments from other laboratories (51, 68) have been taken as evidence that ribosomes can sometimes scan backwards to reinitiate. That interpretation is not supported by other published data (35, 40), however, or by the experiments described herein.

Reinitiation versus direct internal binding. Although the observations reported here and elsewhere (20, 35, 40, 51) are most easily explained by invoking reinitiation, it has not been shown that the putative reinitiation is insensitive to dilution and to inhibitors that prevent de novo ribosome binding. The hypothesis that eucaryotic ribosomes can reinitiate therefore still awaits rigorous proof. However many indirect arguments impugn the alternative idea that ribosomes can initiate directly at internal sites. The primary observation to be explained is that the presence of a terminator codon in frame with the first ATG codon and upstream from the second allows efficient use of the second potential

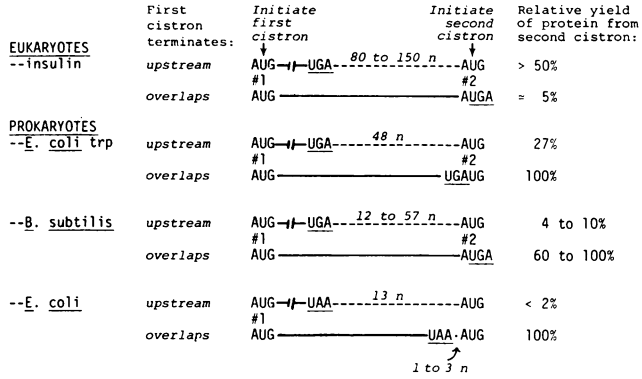

FIG. 6. Consequences of experimentally expanding the intercistronic distance in eucaryotes compared with procaryotes. Data for eucaryotes are from this paper; data for procaryotes are from references 11, 60, and 64. It should be noted that not all overlapping bacterial cistrons show coupled translation, and not all bacterial genes that show coupled expression overlap.

initiation site. One might explain this by postulating that eucaryotic ribosomes can initiate directly at internal ATG codons but fail to do so when the downstream site is occluded by 80S ribosomes advancing from upstream. The evidence against occlusion and in favor of reinitiation includes the following.

(i) With polycistronic procaryotic transcripts, where occlusion can indeed occur, the inhibitory effect of an overlapping upstream cistron is sometimes only two- or threefold (11, 19, 61). In eucaryotes, on the other hand, initiation at a downstream ATG codon is completely suppressed when it is overlapped by an upstream cistron (8, 35, 40, 54, 65). (This is true only under conditions that preclude leaky scanning, and at present it can be shown most reliably in vivo. Cell extracts, especially from reticulocytes, sometimes initiate promiscuously at ATG codons that clearly are not accessible in vivo [26, 45].)

(ii) Eucaryotic ribosomes initiate at the first but not the second ATG codon even when the two ATG triplets are flanked by identical sequences (which precludes favoritism based on context) and are separated by only five nucleotides (see plasmid C2 in reference 35). In that specially designed construct it would seem difficult for elongating 80S ribosomes to selectively block the second ATG codon.

(iii) Occlusion should be relieved when the upstream ORF terminates just before or at the start of the next cistron, yet synthesis of proinsulin by D13A and D15A remained very low (Fig. 3).

(iv) Reinitiation can be blocked by introducing a hairpin structure between the upstream ORF and the preproinsulin start site (unpublished data), a result that seems incompatible with the occlusion argument.

(v) The occlusion model requires direct binding of 40S subunits to internal ATG codons, but that postulate is contradicted by the inability of eucaryotic ribosomes to bind circular mRNAs (27, 28); the ability of duplex RNA-RNA or RNA-DNA structures, located far upstream from the ATG codon, to block translation (37); the trapping of 40S subunits upstream from the ATG codon when ATP reserves are depleted during cell-free translation (29); and the restriction of initiation to the 5′-proximal copy in a tandem array of identical "ribosome binding sites" (32). Various other observations that were once taken as evidence for direct

internal initiation (46, 57, 70) have been reinterpreted recently in the light of new data (4, 7, 13, 55).

Other examples of reinitiation in eucaryotes. The presence of (mini)cistrons in the leader regions of some natural eucaryotic mRNAs makes it likely that reinitiation is not limited to laboratory-manipulated mRNAs introduced into COS cells. Because the motif is not widespread, however (nearly half of the mRNAs that are suspected to support reinitiation are derived from protooncogenes [38]), reinitiation in eucaryotic cells at large might be less efficient than in cultured COS cells. With several bicistronic viral mRNAs, the (poly)peptide encoded in the upstream (mini)cistron has been detected (17, 22, 25, 69), confirming that the leader sequence is indeed translated. In a few cellular mRNAs, the upstream ORF terminates very close to the ATG codon that initiates the major ORF (10, 24, 42, 59, 66); I would expect such mRNAs to reinitiate inefficiently, by analogy with the constructs described in Fig. 3. The demonstration in Fig. 5 that a far-upstream ORF can relieve the potentially inhibitory effect of another upstream ATG codon might also be extrapolated to the handful of naturally occurring mRNAs that are structurally analogous to p255/11BA (3, 5, 14, 44, 48, 63).

Although the efficiency of reinitiation increased dramatically as a function of intercistronic length in our preproinsulin constructs, the generality of that result remains to be seen. Perez et al. (52) have reported efficient reinitiation with chimeric transcripts of simian virus 40 and Rous sarcoma virus that have an intercistronic gap of only 9 nucleotides. On the other hand, the EBNA2 protein is expressed inefficiently despite a lengthy intercistronic gap between the leader ORF and the EBNA2 coding sequence (58, 69). Reinitiation is believed to be the natural mode of translation of caulimovirus genes (12), which are closely spaced (21); thus the requirements for reinitiation in plants might differ from those in mammals.

Reinitiation has not yet been demonstrated in *Saccharomyces cerevisiae* by showing that downstream translation is abolished when the terminator codon between adjacent cistrons is mutated, nor do we know how intercistronic distance affects reinitiation (if it occurs) in yeast. Sherman and Stewart (62) concluded from their study of cytochrome *c* mutants that yeast ribosomes cannot reinitiate, but in all of their mutants the upstream minicistron terminated very close to the ATG triplet that would have been used to reinitiate cytochrome *c*. That motif might be unfavorable for reinitiation in yeast, as it is in higher eucaryotes. Since upstream ATG codons, invariably followed by a terminator codon, have been found in a number of yeast mRNAs (18, 24, 42, 50, 56, 66, 67), it seems likely that yeast ribosomes either reinitiate or undergo leaky scanning, but neither mechanism has yet been established experimentally in yeast.

A practical consequence of reinitiation merits emphasis, namely, that introducing a terminator codon within the body of a gene is not a foolproof way to abolish gene function. The ability of ribosomes to reinitiate downstream from a terminator codon has indeed frustrated some attempts to construct null mutants (1).

Reinitiation in eucaryotes versus procaryotes. The correlation between intercistronic length and the efficiency of reinitiation in eucaryotes is curiously just the opposite of what occurs in procaryotes. In many, albeit not all (2, 47), coregulated bacterial genes, the terminator codon of one cistron overlaps the initiator codon of the next (6, 9, 15, 23, 43, 49, 53, 72, 73) and translation of the downstream cistron is drastically impaired when the intercistronic distance is expanded experimentally (Fig. 6). Perhaps the opposite consequences of increasing intercistronic length are not surprising in view of the many other differences between procaryotes and eucaryotes in mRNA structure and the mechanism of initiation of translation. An important distinction is that eucaryotic ribosomes seem constrained to enter only at the 5′ end of the mRNA, and therefore the same 40S ribosomal subunit presumably translates the first and second cistrons. That is not necessarily true in procaryotes; the fact that the downstream cistron almost always has its own Shine-Dalgarno sequence means that bacterial ribosomes could initiate there de novo; in cases where the distal cistron is translated more efficiently than the cistron that precedes it, ribosomes must initiate de novo at the downstream site (2). The arbitrary order of binding of Met-tRNA and mRNA to bacterial ribosomes also differs from that in eucaryotes, where the small ribosomal subunit must bind Met-tRNA before the ATG initiator codon can be recognized (reviewed in reference 31). If it is the lack of Met-tRNA that makes reinitiation inefficient in eucaryotes when the upstream ORF terminates too close to the next potential start site, that constraint would not extend to procaryotes.

ACKNOWLEDGMENT

Research funds were provided by Public Health Service grant GM 33915 from the National Institutes of Health.

LITERATURE CITED

1. **Barker, D. D., and A. J. Berk.** 1987. Adenovirus proteins from both E1B reading frames are required for transformation of rodent cells by viral infection and DNA transfection. Virology **156**:107–121.
2. **Baughman, G., and M. Nomura.** 1983. Localization of the target site for translational regulation of the L11 operon and direct evidence for translational coupling in Escherichia coli. Cell **34**:979–988.
3. **Bayney, R. M., J. Rodkey, C. Bennett, A. Lu, and C. Pickett.** 1987. Rat liver NAD(P)H:quinone reductase nucleotide sequence analysis. J. Biol. Chem. **262**:572–575.
4. **Bendena, W., M. Abouhaidar, and G. Mackie.** 1985. Synthesis in vitro of the coat protein of papaya mosaic virus. Virology **140**: 257–268.
5. **Betsholtz, C., A. Johnsson, C-H. Heldin, B. Westermark, P. Lind, M. Urdea, R. Eddy, T. Shows, K. Philpott, A. Mellor, T. Knott, and J. Scott.** 1986. cDNA sequence and chromosomal localization of human platelet-derived growth factor A-chain. Nature (London) **320**:695–699.
6. **Blanchin-Roland, S., S. Blanquet, J.-M. Schmitter, and G. Fayat.** 1986. The gene for *E. coli* diadenosine tetraphosphatase is located immediately clockwise to *folA* and forms an operon with *ksgA*. Mol. Gen. Genet. **205**:515–522.
7. **Carrington, J. C., and T. J. Morris.** 1985. Characterization of the cell-free translation products of carnation mottle virus. Virology **144**:1–10.
8. **Cheng, K.-C., G. Smith, and B. Moss.** 1986. Hepatitis B virus large surface protein is not secreted but is immunogenic when selectively expressed by recombinant vaccinia virus. J. Virol. **60**:337–344.
9. **Cole, S. T., T. Grundstrom, B. Jaurin, J. Robinson, and J. Weiner.** 1982. Location and nucleotide sequence of *frdB* encoding the fumarate reductase of *E. coli*. Eur. J. Biochem. **126**: 211–216.
10. **Coussens, L., T. Yang-Feng, Y.-C. Liao, E. Chen, A. Gray, J. McGrath, P. Seeburg, T. Libermann, J. Schlessinger, U. Francke, A. Levinson, and A. Ullrich.** 1985. Tyrosine kinase receptor with extensive homology to EGF receptor shares chromosomal location with *neu* oncogene. Science **230**:1132–1139.

11. **Das, A., and C. Yanofsky.** 1984. A ribosome binding site sequence is necessary for efficient expression of the distal gene of a translationally-coupled gene pair. Nucleic Acids Res. **12:** 4757–4768.
12. **Dixon, L. K., and T. Hohn.** 1984. Initiation of translation of the cauliflower mosaic virus genome from a polycistronic mRNA: evidence from deletion mutagenesis. EMBO J. **3:**2731–2736.
13. **Duncan, R., and P. Dobos.** 1986. The nucleotide sequence of infectious pancreatic necrosis virus dsRNA segment A reveals one large ORF encoding a precursor polyprotein. Nucleic Acids Res. **14:**5934.
14. **Eveleth, D., and J. L. Marsh.** 1986. Sequence and expression of the Cc gene, a member of the dopa decarboxylase gene cluster of *Drosophila*: possible translational regulation. Nucleic Acids Res. **14:**6169–6183.
15. **Finch, P., A. Storey, K. Chapman, K. Brown, I. Hickson, and P. Emmerson.** 1986. Complete nucleotide sequence of the *E. coli recB* gene. Nucleic Acids Res. **14:**8573–8582.
16. **Gluzman, Y.** 1981. SV40-transformed simian cells support the replication of early SV40 mutants. Cell **23:**175–182.
17. **Hackett, P., R. Petersen, C. Hensel, F. Albericio, S. Gunderson, A. Palmenberg, and G. Barany.** 1986. Synthesis in vitro of a seven amino acid peptide encoded in the leader RNA of Rous sarcoma virus. J. Mol. Biol. **190:**45–57.
18. **Hinnebusch, A. G.** 1984. Evidence for translational regulation of the activator of general amino acid control in yeast. Proc. Natl. Acad. Sci. USA **81:**6442–6446.
19. **Hoess, R., C. Foeller, K. Bidwell, and A. Landy.** 1980. Site-specific recombination functions of bacteriophage λ: DNA sequence of regulatory regions and overlapping structural genes for Int and Xis. Proc. Natl. Acad. Sci. USA **77:**2482–2486.
20. **Hughes, S., K. Mellstrom, E. Kosik, F. Tamanoi, and J. Brugge.** 1984. Mutation of a termination codon affects *src* initiation. Mol. Cell. Biol. **4:**1738–1746.
21. **Hull, R., J. Sadler, and M. Longstaff.** 1986. The sequence of carnation etched ring virus DNA: comparison with cauliflower mosaic virus and retroviruses. EMBO J. **5:**3083–3090.
22. **Jay, G., S. Nomura, C. W. Anderson, and G. Khoury.** 1981. Identification of the SV40 agnogene product: a DNA binding protein. Nature (London) **291:**346–349.
23. **Jefferson, R., S. Burgess, and D. Hirsh.** 1986. β-Glucuronidase from *E. coli* as a gene-fusion marker. Proc. Natl. Acad. Sci. USA **83:**8447–8451.
24. **Kammerer, B., A. Guyonvarch, and J. C. Hubert.** 1984. Yeast regulatory gene PPR1. J. Mol. Biol. **180:**239–250.
25. **Khalili, K., J. Brady, and G. Khoury.** 1987. Translational regulation of SV40 early mRNA defines a new viral protein. Cell **48:**639–645.
26. **Koller, K., and M. J. Brownstein.** 1987. Use of a cDNA clone to identify a supposed precursor protein containing valosin. Nature (London) **325:**542–545.
27. **Konarska, M., W. Filipowicz, H. Domdey, and H. Gross.** 1981. Binding of ribosomes to linear and circular forms of the 5'-terminal leader fragment of tobacco mosaic virus RNA. Eur. J. Biochem. **114:**221–227.
28. **Kozak, M.** 1979. Inability of circular mRNA to attach to eukaryotic ribosomes. Nature (London) **280:**82–85.
29. **Kozak, M.** 1980. Role of ATP in binding and migration of 40S ribosomal subunits. Cell **22:**459–467.
30. **Kozak, M.** 1981. Possible role of flanking nucleotides in recognition of the AUG initiator codon by eukaryotic ribosomes. Nucleic Acids Res. **9:**5233–5252.
31. **Kozak, M.** 1983. Comparison of initiation of protein synthesis in procaryotes, eucaryotes, and organelles. Microbiol. Rev. **47:** 1–45.
32. **Kozak, M.** 1983. Translation of insulin-related polypeptides from mRNAs with tandemly reiterated copies of the ribosome binding site. Cell **34:**971–978.
33. **Kozak, M.** 1984. Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. Nucleic Acids Res. **12:**857–872.
34. **Kozak, M.** 1984. Point mutations close to the AUG initiator codon affect the efficiency of translation of rat preproinsulin *in vivo*. Nature (London) **308:**241–246.
35. **Kozak, M.** 1984. Selection of initiation sites by eucaryotic ribosomes: effect of inserting AUG triplets upstream from the coding sequence for preproinsulin. Nucleic Acids. Res. **12:** 3873–3893.
36. **Kozak, M.** 1986. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. Cell **44:**283–292.
37. **Kozak, M.** 1986. Influences of mRNA secondary structure on initiation by eukaryotic ribosomes. Proc. Natl. Acad. Sci. USA **83:**2850–2854.
38. **Kozak, M.** 1986. Bifunctional messenger RNAs in eukaryotes. Cell **47:**481–483.
39. **Kozak, M.** 1986. Regulation of protein synthesis in virus-infected animal cells. Adv. Virus Res. **31:**229–292.
40. **Liu, C.-C., C. Simonsen, and A. D. Levinson.** 1984. Initiation of translation at internal AUG codons in mammalian cells. Nature (London) **309:**82–85.
41. **Lomedico, P. T., and S. J. McAndrew.** 1982. Eukaryotic ribosomes can recognize preproinsulin initiation codons irrespective of their position relative to the 5'-end of mRNA. Nature (London) **299:**221–226.
42. **Losson, R., R. Fuchs, and F. Lacroute.** 1983. *In vivo* transcription of a eukaryotic regulatory gene. EMBO J. **2:**2179–2184.
43. **March, P., J. Ahnn, and M. Inouye.** 1985. The DNA sequence of the gene (*rnc*) encoding ribonuclease III of *E. coli*. Nucleic Acids Res. **13:**4677–4685.
44. **Mardon, G., and H. E. Varmus.** 1983. Frameshift and intragenic supressor mutations in a Rous sarcoma provirus suggest *src* encodes two proteins. Cell **32:**871–879.
45. **Matlashewski, G. J., S. Tuck, D. Pim, P. Lamb, J. Schneider, and L. V. Crawford.** 1987. Primary structure polymorphism at amino acid residue 72 of human p53. Mol. Cell. Biol. **7:**961–963.
46. **Mertens, P. P. C., and P. Dobos.** 1982. Messenger RNA of infectious pancreatic necrosis virus is polycistronic. Nature (London) **297:**243–246.
47. **Mooi, F. R., I. Claassen, D. Bakker, H. Kuipers, and F. K. de Graaf.** 1986. Regulation and structure of an *E. coli* gene coding for an outer membrane protein involved in export of K88ab fimbrial subunits. Nucleic Acids Res. **14:**2443–2457.
48. **Noda, M., T. Ikeda, T. Kayano, H. Suzuki, H. Takeshima, M. Kurasaki, H. Takahashi, and S. Numa.** 1986. Existence of distinct sodium channel messenger RNAs in rat brain. Nature (London) **320:**188–192.
49. **Nohno, T., T. Saito, and J. Hong.** 1986. Cloning and complete nucleotide sequence of the *E. coli* glutamine permease operon *glnHPQ*. Mol. Gen. Genet. **205:**260–269.
50. **Nyunoya, H., and C. J. Lusty.** 1984. Sequence of the small subunit of yeast carbamyl phosphate synthetase and identification of its catalytic domain. J. Biol. Chem. **259:**9790–9798.
51. **Peabody, D. S., S. Subramani, and P. Berg.** 1986. Effect of upstream reading frames on translation efficiency in SV40 recombinants. Mol. Cell. Biol. **6:**2704–2711.
52. **Perez, L., J. W. Wills, and E. Hunter.** 1987. Expression of the Rous sarcoma virus *env* gene from an SV40 late-region replacement vector: effects of upstream initiation codons. J. Virol. **61:**1276–1281.
53. **Perry, K. L., S. J. Elledge, B. Mitchell, L. Marsh, and G. C. Walker.** 1985. *umuDC* and *mucAB* operons whose products are required for UV light- and chemical-induced mutagenesis. Proc. Natl. Acad. Sci. USA **82:**4331–4335.
54. **Persing, D. H., H. E. Varmus, and D. Ganem.** 1986. Inhibition of secretion of hepatitis B surface antigen by a related presurface polypeptide. Science **234:**1388–1391.
55. **Rice, C. M., E. Lenches, S. Eddy, S. J. Shin, R. Sheets, and J. H. Strauss.** 1985. Nucleotide sequence of yellow fever virus: implications for flavivirus gene expression and evolution. Science **229:**726–735.
56. **Rose, M., P. Grisafi, and D. Botstein.** 1984. Structure and function of the yeast *URA3* gene: expression in *E. coli*. Gene **29:113–124.**
57. **Salomon, R., M. Bar-Joseph, H. Soreq, I. Gozes, and U. Z.**

Littauer. 1978. Translation in vitro of carnation mottle virus RNA. Virology **90**:288–298.

58. **Sample, J., M. Hummel, D. Braun, M. Birkenbach, and E. Kieff.** 1986. Nucleotide sequences of mRNAs encoding Epstein-Barr virus nuclear proteins: a probable transcriptional initiation site. Proc. Natl. Acad. Sci. USA **83**:5096–5100.
59. **Schneider, C., M. J. Owen, D. Banville, and J. G. Williams.** 1984. Primary structure of human transferrin receptor deduced from the mRNA sequence. Nature (London) **311**:675–678.
60. **Schoner, B. E., R. M. Belagaje, and R. G. Schoner.** 1986. Translation of a synthetic two-cistron mRNA in *E. coli*. Proc. Natl. Acad. Sci. USA **83**:8506–8510.
61. **Schumperli, D., K. McKenney, D. A. Sobieski, and M. Rosenberg.** 1982. Translational coupling at an intercistronic boundary of the *E. coli* galactose operon. Cell **30**:865–871.
62. **Sherman, F., and J. W. Stewart.** 1982. Mutations altering initiation of translation of yeast iso-1-cytochrome c; contrasts between the eukaryotic and prokaryotic initiation process, p. 301–333. *In* J. N. Strathern, E. W. Jones, and J. R. Broach (ed.), Molecular biology of the yeast saccharomyces: metabolism and gene expression. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
63. **Skern, T., W. Sommergruber, D. Blaas, P. Gruendler, F. Fraundorfer, C. Pieler, I. Fogy, and E. Kuechler.** 1985. Human rhinovirus 2: complete nucleotide sequence and proteolytic signals in the capsid protein region. Nucleic Acids Res. **13**: 2111–2126.
64. **Sprengel, R., B. Reiss, and H. Schaller.** 1985. Translationally coupled initiation of protein synthesis in *Bacillus subtilis*. Nucleic Acids Res. **13**:893–909.
65. **Standring, D. N., J.-H. Ou, and W. J. Rutter.** 1986. Assembly of viral particles in *Xenopus* oocytes: pre-surface antigens regulate secretion of the hepatitis B viral surface envelope particle. Proc. Natl. Acad. Sci. USA **83**:9338–9342.
66. **Sutton, A., and J. R. Broach.** 1985. Signals for transcription initiation and termination in the *Saccharomyces cerevisiae* plasmid 2μm circle. Mol. Cell. Biol. **5**:2770–2780.
67. **Thireos, G., M. D. Penn, and H. Greer.** 1984. 5′ untranslated sequences are required for the translational control of a yeast regulatory gene. Proc. Natl. Acad. Sci. USA **81**:5096–5100.
68. **Thomas, K. R., and M. R. Capecchi.** 1986. Introduction of homologous DNA sequences into mammalian cells induces mutations in the cognate gene. Nature (London) **324**:34–38.
69. **Wang, F., L. Petti, D. Braun, S. Seung, and E. Kieff.** 1987. A bicistronic Epstein-Barr virus mRNA encodes two nuclear proteins in latently infected, growth-transformed lymphocytes. J. Virol. **61**:945–954.
70. **Westaway, E. G.** 1977. Strategy of the flavivirus genome: evidence for multiple internal initiation of translation of proteins specified by Kunjin virus in mammalian cells. Virology **80**:320–335.
71. **Wigler, M., A. Pellicer, S. Silverstein, and R. Axel.** 1978. Biochemical transfer of single-copy eucaryotic genes using total cellular DNA as donor. Cell **14**:725–731.
72. **Yamamoto, T., T. Tamura, T. Yokota, and T. Takano.** 1982. Overlapping genes in the heat-labile enterotoxin operon originating from *E. coli* human strain. Mol. Gen. Genet. **188**: 356–359.
73. **Yanofsky, C., T. Platt., I. P. Crawford, B. P. Nichols, G. E. Christie, H. Horowitz, M. VanCleemput, and A. M. Wu.** 1981. The complete nucleotide sequence of the tryptophan operon of *Escherichia coli*. Nucleic Acids Res. **9**:6647–6667.