Circumstances and Mechanisms of Inhibition of Translation by Secondary Structure in Eucaryotic mRNAs

MARILYN KOZAK

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

Received 23 June 1989/Accepted 26 July 1989

This paper describes *in vitro* experiments with two types of intramolecular duplex structures that inhibit translation *in cis* by preventing the formation of an initiation complex or by causing the complex to be abortive. One stem-loop structure ($\Delta G = -30$ kcal/mol) prevented mRNA from engaging 40S subunits when the hairpin occurred 12 nucleotides (nt) from the cap but had no deleterious effect when it was repositioned 52 nt from the cap. This result confirms prior *in vivo* evidence that the 40S subunit-factor complex, once bound to mRNA, has considerable ability to penetrate secondary structure. Consequently, translation is most sensitive to secondary structure at the entry site for ribosomes, i.e., the 5' end of the mRNA. The second stem-loop structure (hp7; $\Delta G = -61$ kcal/mol, located 72 nt from the cap) was too stable to be unwound by 40S ribosomes. hp7 did not prevent a 40S ribosomal subunit from binding but caused the 40S subunit to stall on the 5' side of the hairpin, exactly as the scanning model predicts. Control experiments revealed that 80S elongating ribosomes could disrupt duplex structures, such as hp7, that were too stable to be penetrated by the scanning 40S ribosome-factor complex. A third type of base-paired structure shown to inhibit translation *in vivo* involves a long-range interaction between the 5' and 3' noncoding sequences.

The scanning model for initiation in eucaryotes postulates three steps in the association of mRNA with 40S ribosomal subunits: binding of a 40S ribosomal subunit-factor complex to the capped 5' end of the mRNA (step 1), followed by linear migration (step 2) which ends when the 40S subunit reaches (usually) the first AUG codon (step 3). Step 1 is facilitated by the m⁷G cap (29) and a well-characterized cap-binding protein (30). Step 3, recognition of the AUG codon, is mediated by eIF2-modulated base pairing with the anticodon in Met-tRNA^{Met} (5, 6) and, at least in higher eucaryotes, by a consensus sequence that flanks the AUG codon (12, 14, 15). Step 2 is supported by evidence that is provocative but indirect, namely, the effects of the antibiotic edeine on formation of initiation complexes (22) and the strong, predictable effects of position on selection of the initiator codon (5, 10; reviewed in reference 17). Here I present additional evidence for the scanning step by analyzing the binding of wheat germ ribosomes to a transcript that has a very stable stem-loop structure between the cap and the first AUG codon. I show that 40S ribosomal subunits bind with high efficiency to such transcripts but cannot advance to the AUG codon. Instead, a 40S subunit stalls on the 5' side of the hairpin. Thus, inserting a stable stem-loop structure into the mRNA leader sequence provides a way to trap an elusive intermediate in the initiation process.

The aforementioned finding emerged in the course of a broader study of the circumstances and *cis*-limited mechanisms of inhibition of translation by mRNA secondary structure. There have been numerous reports concerning the ability of secondary structure to inhibit translation of eucaryotic mRNAs *in vivo* (reviewed in reference 17; see also Discussion). Here I have resorted to cell-free translation systems to study the level at which translation is blocked by two specific structures: (i) a moderately stable hairpin ($\Delta G = -30$ kcal/mol) which does not inhibit translation when it occurs around the AUG initiator codon (i.e., some distance downstream from the cap) but does inhibit when it is moved close to the 5' end of the transcript and (ii) a very stable hairpin ($\Delta G = -61$ kcal/mol) which inhibits, no matter where

it occurs within the 5' noncoding region, by halting the migration of 40S ribosomal subunits. I also describe circumstances under which long-range base pairing, which involves the 5' and 3' noncoding regions of mRNA, inhibits translation *in vivo*.

MATERIALS AND METHODS

Construction of plasmids. SP64-based plasmids that have near the 5' end a hairpin of -30 kcal/mol (calculated as described in reference 34) were obtained by subcloning into an SP64 · CAT vector the small *Hind*III-*Bam*HI fragment from B13hp, which is a simian virus 40 (SV40)-based preproinsulin-encoding plasmid described previously (13). SP64 · B13hp(12)CAT, in which the hairpin occurs 12 nucleotides (nt) down from the cap, was obtained by inserting the 38-base-pair (bp) hairpin-containing *Hind*III-*Bam*HI fragment from SV40 · B13hp in place of the corresponding portion of the SP64 polylinker; the relevant sequences are given in Fig. 1. To obtain SP64 · B13hp(52)CAT, in which the hairpin occurs 52 nt down from the cap, the aforementioned 38-bp fragment was ligated to an adapter (GATCCTCTGCAGA
GAGACGTCTTCA), and the resulting 51-bp hairpin-containing fragment was inserted at the *Bam*HI site of SP64 · CAT (Fig. 1, line 2). The sequences thus introduced into SP64 · B13hp(12)CAT and SP64 · B13hp(52)CAT include an ATG codon (Fig. 1). To move that ATG codon in frame with the chloramphenicol acetyltransferase (CAT)-coding sequence, an adapter (oligonucleotide 8757) was inserted between the hairpin and the start of the CAT open reading frame. (Oligonucleotide 8757 includes an additional out-of-frame ATG codon to ensure that no ribosomes initiate at the authentic CAT start site; i.e., any CAT-related proteins produced by the B13hp derivatives must initiate at the first ATG codon [the one shown in boldface in Fig. 1]). The protocol for constructing the control plasmid SP64 · B13 · CAT (Fig. 1, line 3) was identical to that used for SP64 · B13hp(12)CAT except that SV40 · B13 (described in references 12 and 13), rather than SV40 · B13hp, was the source of the *Hind*III-*Bam*HI fragment.

FIG. 1. Derivation of SP6-based transcripts that contain a stem-loop structure with a calculated stability of -30 kcal/mol. Sequences are shown with T in place of U. The *Hind*III-to-*Bam*HI portion of the polylinker in the parental construct SP64(8757)CAT (line 4) was replaced by a 21-bp ATG-containing deoxyoligonucleotide (uppercase letters, line 3) to obtain SP64 · B13hp(12)CAT. SP64 · B13hp(52)CAT was constructed by inserting a 51-bp ATG-containing deoxyoligonucleotide (uppercase letters, line 2) at the *Bam*HI site of SP64(8757)CAT. All four constructs are identical downstream from the adapter (oligonucleotide 8757), which is overlined. Initiation at the ATG codon shown in boldface in the three B13 derivatives produces a CAT-related protein (pre-CAT) with an extra 19 amino acids at the amino terminus. The natural ATG start site for CAT (underlined in SP64 · B13 · CAT) is silent in these constructs.

An SP64-based plasmid that has near its 5' end a very stable hairpin ($\Delta G = -61$ kcal/mol) was constructed in three steps. First I made SP64 · L by inserting into SP64(8180)_{x3}CAT, in place of the *Hind*III-to-*Bam*HI portion of the polylinker, the sequence AGCTTGGGGCGCGTGG TGGCGGCTGCAG. In a separate construction, I inserted the complementary sequence (see Fig. 5) to obtain SP64 · R. SP64 · L and SP64 · R were designed to serve as controls, since they have no deliberate secondary structure in the leader sequence. SP64 · L and SP64 · R were so named because, when they were recombined via the unique *Pst*I sites (step 3 in the construction), the oligonucleotide inserts in L and R formed the left and right sides of the stem in hp7 (see Fig. 5).

SV40-based plasmids that encode preproinsulin were manipulated to investigate the in vivo consequences of long-range interactions in eucaryotic mRNAs. Oligonucleotide 8153, which is complementary to a sequence that occurs naturally between the terminator codon and polyadenylation signal (AATAAA) of the rat preproinsulin II gene (31), was inserted at the *Hind*III site near the 5' end of the standard SV40-preproinsulin construct p255/11 (11). The result was a plasmid (Fig. 2, construct 1) in which the 5' and 3' noncoding sequences are complementary over 18 nt. As a control, oligonucleotide 8232, which has a 4-bp mismatch relative to the target 3' noncoding sequence, was inserted at the same 5' location (Fig. 2B, construct 2). Oligonucleotide 8208, which is similar to 8153 except for having a *Bam*HI rather than a *Hind*III-complementary 5' terminus, was inserted at the *Bam*HI site near the start of the signal peptide-coding sequence (Fig. 2B, construct 4). Finally, to investigate a long-range interaction between the 5' noncoding and part of the preproinsulin-coding sequence, oligonucleotide 8210, which is complementary to a sequence at the junction of the C and A domains of preproinsulin, was inserted at the 5'-proximal *Hind*III site of p255/11 (Fig. 2B, construct 3).

The structures of all of these plasmids were confirmed by dideoxy-chain termination sequencing reactions, using appropriately positioned primers.

Transfection of COS cells for in vivo translation experiments. The transfection technique has been described (10, 12), along with procedures for labeling, extracting, and analyzing proinsulin. Cytoplasmic RNA levels were monitored by dot blot hybridization and were found not to vary among the plasmids studied.

In vitro transcription and translation protocols. Transcription with SP6 polymerase (Pharmacia, Inc.) was performed as described previously (18), using DNA templates linearized with *Ava*I. For translation in rabbit reticulocyte lysates (Bethesda Research Laboratories), the Mg^{2+} and potassium concentrations were 2.2 and 135 mM, respectively. The mRNA concentration was subsaturating at ~ 0.1 μ g per 25- μ l reaction. Initiation complexes formed by incubating capped radiolabeled mRNAs with wheat germ S23 extracts were analyzed according to published protocols (19) except that the Mg^{2+} concentration was lowered to 1.2 mM to improve the fidelity of initiation (18). For the analysis of initiation complexes, polypeptide elongation was inhibited by including 200 μ M sparsomycin, which was a gift from the Division of Cancer Treatment of the National Cancer Institute. Formation of 40S and 80S initiation complexes was monitored by sedimentation through 10 to 30% glycerol gradients containing 20 mM *N*-2-hydroxyethylpiperazine-*N'*-2-ethanesulfonic acid (HEPES; pH 7.5), 75 mM KCl, 1.5 mM magnesium acetate, and 0.1 mM EDTA. The gradients were centrifuged at 4°C for 4 h at 39,000 rpm in a Beckman SW41 rotor. Fractions from analytical gradients were mixed with Aquasol-2 (Du Pont Co.) for counting.

Fingerprinting of ribosome-protected mRNA fragments. Ribosome-protected fragments were obtained by incubating 32 P-labeled SP64 · hp7 mRNA, or similarly labeled control transcripts, with wheat germ ribosomes for 10 min in the

FIG. 2. Test for long-distance base-pairing effects on the translation of preproinsulin. COS cells were transfected with the wild-type plasmid p255/11 or with one of the insertion mutants listed in panel B. Details of the constructions are given in Materials and Methods. The diagram in part A shows two naturally occurring sequences, one in the coding region (designated CA) and one in the 3' untranslated region, to which the oligonucleotides shown in panel B are complementary. (C and D) Cropped fluorograms of [³⁵S]cysteine-labeled, immunoprecipitated proinsulin synthesized by 2-day transfected cells during a 2-h labeling period in the presence or absence of excess NaCl, as indicated. Protein yields were quantitated by densitometry (12). The plasmids used for transfection are identified above each three-lane block of the polyacrylamide gel.

presence of sparsomycin, as described above. The initiation complexes were digested with pancreatic RNase (10 µg/ml, 10 min, 20°C) before centrifugation through glycerol gradients. The conditions used for binding mRNA, digesting, and analyzing protected fragments were similar to techniques worked out for reovirus mRNAs (20, 21). Briefly, nuclease-resistant ³²P-labeled fragments recovered from the 40S and 80S regions of the gradients were purified by phenol extraction, followed by electrophoresis through a 20% polyacrylamide-8 M urea gel. After elution from the gel, the fragments were digested to completion with T₁ RNase (5,000 U/ml, 30 min, 37°C) or pancreatic RNase (800 µg/ml). The resulting oligonucleotides were fractionated by high-voltage electrophoresis on cellulose acetate strips, followed by homochromatography (see legend to Fig. 6). T₁ and pancreatic RNase oligonucleotides eluted from thin-layer plates were identified by secondary digestion with pancreatic RNase (for T₁ fingerprints) or T₁ RNase (for pancreatic fingerprints). This was done on three RNA preparations labeled, respectively, with [α-³²P]GTP, [α-³²P]ATP, and [α-³²P]CTP. Autoradiograms were exposed with intensifying screens at -70°C.

RESULTS

Position-dependent inhibition by a moderately stable hairpin structure. If the scanning 40S subunit-factor complex,

FIG. 3. Translation of secondary structure-containing mRNAs in a rabbit reticulocyte translation system. A cropped autoradiogram of [³⁵S]methionine-labeled pre-CAT proteins, analyzed by polyacrylamide gel electrophoresis, is shown. The SP6-derived mRNAs used for translation in lanes 1 to 4 are identified in Fig. 1, and the sequence of SP64 · hp7 (lane 5) is given in Fig. 5. For lanes 4 and 6, each of the two mRNAs was used at the same concentration as in the corresponding single-mRNA lanes.

once bound to mRNA, has some capacity to unwind the mRNA, a reasonable prediction is that secondary structure should inhibit translation more effectively when it occurs at the entry site for ribosomes (i.e., the 5' end of the mRNA) than when secondary structure occurs internally. To test that hypothesis, the same hairpin structure was introduced in two different positions within the 5' noncoding region of SP64 · CAT transcripts. These transcripts, SP64 · B13hp(12)CAT and SP64 · B13hp(52)CAT, along with the structure-free control transcript SP64 · B13 · CAT, are depicted in Fig. 1. Their translation in a reticulocyte lysate is shown in Fig. 3. Comparison of lanes 1 and 2 shows that a hairpin with a calculated stability of -30 kcal/mol did not reduce the yield of CAT protein when the hairpin was positioned 52 nt from the 5' end of the transcript, consistent with earlier *in vivo* findings (13). The yield of CAT protein was reduced by 75%, however, when the same hairpin was located 12 nt from the cap (Fig. 3, lane 3). The observed inhibition was *cis* limited inasmuch as the yield of CAT from SP64 · B13 · CAT was not reduced when equal amounts of that transcript and B13hp(12)CAT were cotranslated (Fig. 3, lane 4).

At the primary sequence level, the functionally impaired SP64 · B13hp(12)CAT transcript has a 17-nt insertion (AGCTTGGGCCGTGGTGG) that is absent from the otherwise identical SP64 · B13 · CAT transcript. Another control construct, SP64 · B35(17a)CAT [adapted from SV40 · B35(17a) described in reference 13], which has the same 17-nt insert as SP64 · B13hp(12)CAT but a different sequence immediately downstream, showed no translational dysfunction (M. Kozak, unpublished data); therefore, one may conclude that the restricted translation of SP64 · B13hp(12)CAT is due to complementarity between the 17-nt insert and the adjacent downstream sequence (Fig. 1). Genetic evidence thus defines the 5'-proximal hairpin in SP64 · B13hp(12)CAT as the cause of its poor translation.

The difference in function between SP64 · B13hp(12)CAT and SP64 · B13hp(52)CAT, illustrated in Fig. 3 by monitor-

FIG. 4. Influence of mRNA secondary structure on formation of initiation complexes. ^3H -labeled transcripts from SP64 · B13 · CAT (A), SP64 · B13hp(52)CAT (B), SP64 · B13hp(12)CAT (C), and SP64 · hp7 (D) were incubated for 10 min in a wheat germ translation system in the presence of sparsomycin. Initiation complexes formed with full-length mRNAs (i.e., not treated with RNase) were assayed by glycerol gradient centrifugation. Sedimentation was from right to left. The wheat germ system was used here because its capacity for forming initiation complexes is about 10-fold higher than that of standard reticulocyte lysates.

ing CAT synthesis, was reproduced with an assay specific for initiation. Thus, the yields of 80S sparsomycin-blocked complexes were identical with ^3H -labeled SP64 · B13 · CAT and SP64 · B13hp(52)CAT transcripts (Fig. 4A and B), but the formation of 80S initiation complexes was decreased by two-thirds when mRNA from SP64 · B13hp(12)CAT was tested (Fig. 4C). (The reaction conditions used for the reticulocyte and wheat germ translation experiments in Fig. 3 and 4 are conditions that have been shown to support correct, context-dependent recognition of the AUG initiation codon [18].)

Trapping of 40S ribosomal subunits upstream from a hairpin that cannot be melted. In contrast with the -30 -kcal/mol hairpin that did not impair translation when it was located at

the midpoint of the 5' noncoding sequence, previous experiments revealed that a similarly positioned hairpin of -50 kcal/mol profoundly inhibited the translation of preproinsulin in vivo (13). The simplest explanation was that a 40S ribosomal subunit could not scan beyond an unmeltable base-paired structure. As it was difficult to test that interpretation in vivo, I constructed an SP6-based vector with which to study the problem in vitro. The hairpin I designed has a calculated stability of -61 kcal/mol and is located 72 nt from the 5' end of the transcript (SP64 · hp7 in Fig. 5). Consistent with the aforementioned in vivo result, SP64 · hp7 synthesized no detectable CAT protein in vitro (Fig. 3, lane 5). The inhibitory effect was *cis* limited, as shown by cotranslation of equal amounts of mRNA from SP64 · hp7 and

FIG. 5. Structure of an SP6 vector-derived transcript that has a very stable stem-loop structure ($\Delta G = -61$ kcal/mol) upstream from the CAT start site. Sequences are shown with T in place of U. SP64 · hp7 was derived by recombining SP64 · L and SP64 · R through the *Pst*I sites (overlined). When the control transcripts L and R were incubated with wheat germ ribosomes in the presence of sparsomycin, 80S initiation complexes assembled at the ATG codon shown in boldface, which is the start of the CAT-coding sequence. When SP64 · hp7 transcripts were incubated in the same way, 40S ribosomal subunits bound and protected the 5' end of the mRNA up to and including the hairpin. The T1 and pancreatic RNase oligonucleotides identified in fingerprints of ribosome-protected mRNA fragments (see Fig. 6) are marked T and cP (c indicates control), respectively. Oligonucleotides shown in parentheses were recovered in submolar yields, reflecting their locations at the ends of the ribosome-protected fragments.

FIG. 6. Fingerprint analyses of control and ribosome-protected fragments. RNA fragments labeled with [α - 32 P]GTP were digested with T1 (A to D) or pancreatic (E and F) RNase. Oligonucleotides were fractionated in the first dimension (from left to right) by electrophoresis at pH 3.5 on cellulose acetate strips and in the second dimension (from bottom to top) by homochromatography. Polyethyleneimine-cellulose thin-layer plates were used for homochromatography in panels A to D, and DEAE-cellulose plates were used in panels E and F. To facilitate identification of the ribosome-protected RNA fragment from hp7, its fingerprints (C and F) are

SP64 · B13 · CAT (Fig. 3, lane 6). A striking result was obtained when hp7 transcripts were tested in an initiation assay: the mRNA accumulated almost exclusively, and very efficiently, in 40S initiation complexes (Fig. 4D). The restricted joining of 60S subunits is consistent with the finding (described next) that 40S subunits stall on the 5' side of the hairpin; i.e., the hairpin evidently prevents 40S subunits from reaching the downstream AUG codon that would normally trigger 60S joining. (The small amount of hp7 mRNA that sedimented with 80S ribosomes in Fig. 4D yielded, after digestion with RNase, the same fingerprint as 40S ribosome-protected fragments [see below], indicating that no ribosomes reach the AUG initiator codon in this mRNA. A likely explanation for the small peak of 80S complexes in Fig. 4D is that, to some extent, 60S subunits tend to couple, abortively, with stalled 40S subunits even when the latter are not at an AUG codon.)

To pinpoint the position of the stalled 40S ribosome on hp7 mRNA, the 3 H-labeled mRNA used in Fig. 4D was replaced with 32 P-labeled mRNA, and the initiation complexes were trimmed with RNase before gradient fractionation. An analysis of the protected portion of hp7 mRNA recovered from the 40S region of a glycerol gradient is presented in Fig. 6, along with appropriate controls. I will describe the controls first. SP64 · L and SP64 · R served both as precursors for the construction of hp7 (see Materials and Methods) and as controls for ribosome binding. All of the ribosome-protected 32 P-labeled material from SP64 · L and SP64 · R sedimented with 80S ribosomes, and T1 and pancreatic RNase fingerprints revealed that the 80S sparsomycin-blocked ribosome was centered around the AUG initiator codon for CAT. The results are diagrammed in Fig. 5, and a representative T1 RNase fingerprint is shown in Fig. 6D. The ribosome-protected fragment obtained with hp7, on the other hand, gave a fingerprint that corresponded not to the sequence around the AUG codon but to the 5' end of the mRNA. The size of the 40S ribosome-protected material obtained with hp7 was 75 to 80 nt, versus 30 to 35 nt for the 80S ribosome-protected fragments from SP64 · L and SP64 · R. (A similar difference in size between 40S- and 80S-protected fragments was seen in earlier studies with reovirus mRNAs [20, 21]; presumably, it reflects association of a large mass of initiation factors with the 40S ribosomal subunit.) A T1 RNase fingerprint of the ribosome-protected material from hp7 is shown in Fig. 6C, and the oligonucleotides labeled therein are identified in Fig. 5. The identification was accomplished by analysis of secondary digestion products as described in Materials and Methods. For purposes of orientation, I have included in Fig. 6 T1 RNase fingerprints of 5'

compared with those of 5' terminal fragments (not ribosome protected) from SP64 · L (A and E) and SP64 · R (B). Oligonucleotides marked with asterisks in panels A, B, and E derive from the 5' terminus of the mRNA; they were present in the 5' L_{Pst} and R_{Bam} fragments but not in the 40S ribosome-protected material from hp7. The T1 RNase oligonucleotides numbered in A to D are identified in Fig. 5. The pancreatic RNase oligonucleotides displayed in panels E and F are identified in panel G. GGGGC, the largest pancreatic oligonucleotide in panels E and F, derives from near the base of the left side of the stem in hp7. The 40S-protected fingerprint in panel F has only one pancreatic oligonucleotide (GGAU) that is absent from the 5' L_{Pst} fingerprint in E; GGAU flanks the base of the right side of the stem in hp7. Contaminating RNA fragments that adhered to 80S ribosomes, but were not actually protected in the initial binding step, might account for the extraneous spots (di- and trinucleotides) in panel D.

FIG. 7. Structure of SP64(8387)hp7 in which an ATG codon, introduced as part of oligonucleotide 8387, precedes the hairpin structure. In both SP64(8387)hp7 and the unstructured control construct SP64(8387)L, the adventitious upstream ATG codon is in frame with the CAT-coding sequence. The translation of an elongated pre-CAT protein by both constructs is shown in Fig. 8.

fragments (not ribosome protected) from SP64 · L and SP64 · R (Fig. 6A and B, respectively). The main point is that the T1 oligonucleotides from the protected region of hp7 are the sum of the oligonucleotides from the 5' ends of SP64 · L and SP64 · R, lacking only the extreme 5'-terminal oligonucleotide. I conclude that a 40S ribosomal subunit protected the entire sequence upstream from the unmeltable hairpin structure in hp7 mRNA, extending close to but not including the 5' capped terminus. That conclusion is fully supported by the pancreatic RNase fingerprints in Fig. 6E and F.

The foregoing description applies to the major protected product consistently obtained with ³²P-labeled hp7 mRNA. I also recovered, in variable yields, a protected 5' fragment that extended only as far as the loop in hp7, lacking the right side of the stem. It could be that the pancreatic RNase used to trim the initiation complexes hits most often within the hairpin loop but that 40S ribosomal subunits immediately run off the resulting 5' fragment, thus explaining its poor and variable recovery. Although RNase might cut less frequently at the base of the stem, those fragments, from which 40S subunits cannot run off, would be preferentially recovered.

It is clear that 40S ribosomal subunits were trapped upstream because of the secondary structure of SP64 · hp7. Thus, although SP64 · L and SP64 · R transcripts contain sequences that correspond, respectively, to the left and right sides of the stem in hp7, 40S subunits did not form stable complexes near the 5' ends of those control transcripts. (If the concentration of 40S subunits were high enough, extra 40S ribosomes would actually be expected to load onto the long leader sequence of the control transcripts; however, upon cleavage with RNase, there would be nothing to hold the extra 40S subunits in place. Thus, a different assay would have to be devised to test that aspect of the scanning model.)

80S elongating ribosomes can penetrate hp7. The foregoing results established that 40S ribosomal subunits, with associated initiation factors, could not disrupt the stable stem-loop structure that occurred upstream from the AUG codon in SP64 · hp7. To determine whether that structure could be penetrated by 80S elongating ribosomes, an AUG-containing oligonucleotide (oligonucleotide 8387 in Fig. 7) was introduced at the *Hind*III site just 5' to the hairpin. Translation of

the resulting derivative by reticulocyte and wheat germ ribosomes is shown in Fig. 8A and B, respectively. Whereas SP64 · hp7, in which the first AUG codon resides 3' to the hairpin, was incapable of being translated (lane 1), SP64(8387)hp7 directed synthesis of an N-terminally extended CAT protein (pre-CAT, lane 2) that initiated from the adventitious upstream AUG codon. In other words, although a scanning 40S ribosomal subunit could not get through the stem-loop structure in SP64 · hp7, introducing an AUG codon 5' to the hairpin converted the 40S initiation complexes to 80S elongating ribosomes, at least some of which could penetrate the barrier. The yield of pre-CAT protein was understandably lower from SP64(8387)hp7 than from the control plasmid SP64(8387)L, which has no structural barrier (Fig. 8, lanes 4). As predicted by the scanning model, the presence of a strong upstream AUG codon in SP64(8387)L completely supplanted initiation from the normal CAT start site (Fig. 8, lanes 4 versus lanes 3).

Long-range base pairing in eucaryotic mRNAs. The intro-

FIG. 8. Expression of SP64(8387)hp7 and related constructs in rabbit reticulocyte (A) and wheat germ (B) translation systems. The structures of the four plasmids tested are given in Fig. 5 and 7. Proteins were fractionated by polyacrylamide gel electrophoresis as for Fig. 3. [³⁵S]methionine-labeled CAT (◄) and pre-CAT (►) proteins are indicated; the latter initiates from the upstream in-frame ATG (AUG) codon within oligonucleotide 8387.

duction of complementarity between the 5' and 3' noncoding regions of preproinsulin mRNA can inhibit its translation in vivo, as shown with construct 1 in Fig. 2. Translation was not compromised by the mere presence of 5'-3' complementarity; rather, the cells had to be stressed by adding a slight excess of NaCl to the medium. Thus, under isotonic conditions, equally efficient synthesis of proinsulin was observed with wild-type p255/11 and construct 1 (Fig. 2C, lanes 1 and 4); the addition of 100 mM NaCl, however, caused a twofold reduction in translation of construct 1 (Fig. 2D, lane 4 versus lane 5), whereas translation of the wild-type transcript increased twofold (Fig. 2D, lane 1 versus lane 2; see reference 16 for discussion and further documentation of this stimulation). A slightly higher concentration of NaCl (125 mM in Fig. 2D or 115 mM in Fig. 2C), which had no adverse effect on the translation of wild-type p255/11, caused a sixfold reduction in the translation of construct 1. In contrast, synthesis of proinsulin by construct 2 (control in Fig. 2C) was virtually identical to that of the wild-type plasmid at 115 mM NaCl. Construct 2 has a 5' insertion similar to that of construct 1 except that the substitution of 4 C's for 4 G's (Fig. 2B) diminishes the complementarity to the target 3' noncoding sequence.

Whereas complementarity between 5' and 3' noncoding sequences caused translation to be sensitive to hypertonic stress (Fig. 2C, lane 4 versus lane 5), complementarity between the 5' noncoding and part of the preproinsulin-coding sequence had no such effect (Fig. 2C, lane 10 versus lane 11). Similarly, the insertion near the beginning of the coding sequence (i.e., within the signal peptide region) of an oligonucleotide complementary to the 3' noncoding sequence had no deleterious effect on translation (Fig. 2D, lanes 7 to 9 versus lanes 1 to 3).

DISCUSSION

The three types of intramolecular duplex structures studied here reveal *cis*-limited inhibitory effects on initiation of translation that depend, in understandable ways, on the strength and position of the base-paired region.

The first type of stem-loop structure, which has a calculated stability of -30 kcal/mol, impairs translation in vitro only when it is positioned so close to the 5' end of the mRNA (12 nt from the cap) that a 40S ribosomal subunit cannot bind. That result obtained with SP64 · B13hp(12)CAT (Fig. 4C) is consistent with the scanning model, which postulates that the 5' end of the mRNA is the entry site for the 40S subunit-factor complex. The fact that the -30 -kcal/mol hairpin no longer inhibited translation when it was relocated 52 nt from the 5' terminus (Fig. 3, lane 2; Fig. 4B) confirms prior in vivo evidence (13) that the 40S ribosome-factor complex, once bound, has considerable ability to penetrate secondary structure. Effects of a -30 -kcal/mol hairpin in vivo are somewhat more complicated in that they depend on the conditions under which cells are grown (16); that consideration has not yet been followed up in vitro.

The second type of stem-loop structure studied here ($\Delta G = -61$ kcal/mol, located 72 nt from the cap) was designed to find out how ribosomes deal with an mRNA in which the hairpin is too stable to be unwound. A key prediction of the scanning model is that a 40S ribosomal subunit should bind efficiently to mRNA from plasmid SP64 · hp7 but that the 40S ribosome-factor complex should stall on the 5' side of the hairpin. The ribosome protection data summarized in Fig. 5 and 6 exactly confirm those expectations. The inability of a 40S subunit to jump a hairpin that is too stable to be

melted is strong evidence of the linearity of scanning. The absence of initiation from the downstream CAT start site, which is in a favorable context for initiation and is not sequestered by secondary structure, is consistent with the inability of eucaryotic ribosomes to initiate internally except when the mRNA has been broken (reviewed in reference 17). In addition to elucidating the mechanism of inhibition by a particular stem-loop structure, the experiments described here provide a means to accumulate an elusive intermediate in translation, namely, a complex in which the 40S ribosomal subunit has apparently initiated scanning but has not yet reached the AUG codon. It will be of interest to identify the initiation factors associated with the 40S ribosome-mRNA complex at that early step.

The fact that some pre-CAT protein was synthesized when an AUG codon was introduced 5' to hp7 (Fig. 8, lanes 2) implies that that remarkably stable hairpin, which could not be penetrated by 40S initiating ribosomes, was disrupted by at least some 80S ribosomes during the elongation stage of polypeptide synthesis. This raises the possibility that upstream AUG codons, when they are present, actually facilitate the translation of mRNAs with highly structured leader sequences. The greater melting power of 80S elongating ribosomes than of 40S initiation complexes may also explain why complementary (antisense) sequences targeted to portions of the coding sequence sometimes inhibit translation less effectively than do antisense sequences targeted to the 5' noncoding region (24, 35; Fig. 2).

The third type of inhibitory structure demonstrated herein involves a long-range complementary interaction between the 5' and 3' noncoding sequences. Long-range base pairings are known to form in some bacterial transcripts (2, 3, 25), and there is one published report of repression of eucaryotic translation in vitro by pairing between 5' and 3' noncoding sequences (33). Here I show that inhibition of translation by long-range interactions can also be demonstrated in the cytoplasm of transfected cells, where the inhibition depends on the composition of the culture medium and, in an interesting way, on the precise location of the complementary sequences within the mRNA. The fact that translation of construct 1 (Fig. 2) was inhibited only when the concentration of NaCl was elevated might be explained in several ways. It could be that when the translational machinery is not limiting, the 5' leader sequence is always occupied by 40S ribosomal subunits, and thus shielded, the 5' noncoding sequence cannot anneal with its 3' counterpart. Because hypertonic stress reduces the capacity of the cells to initiate translation (23, 28), the competition between ribosomes and the 3' noncoding sequence for binding to the 5' end of the mRNA would be shifted in favor of 3'-5' pairing. An alternative explanation is that transient exposure to high salt might displace some of the ubiquitous ribonucleoproteins, and the resulting exposure of 5' and 3' noncoding sequences would allow them to anneal. The fact that translation was not impaired when complementarity was introduced in a way that involved the coding, rather than exclusively the noncoding, regions of preproinsulin mRNAs (constructs 3 and 4 in Fig. 2) might be attributed to the greater melting power of 80S ribosomes, as diagrammed in Fig. 9. While further experiments are needed to evaluate that explanation, the data in Fig. 2 tell us at least that, in addition to local secondary structures, long-range interactions in eucaryotic mRNAs can affect translation in vivo. Whether such interactions have regulatory significance for the expression of natural mRNAs remains to be seen. There is some electron microscopic evidence for circular mRNAs in cells (4, 8). The

FIG. 9. Model to explain why complementary sequences impair translation only when the base pairing is restricted to noncoding sequences. (A) Preproinsulin-coding sequence (■); (B) steady-state loading achieved under conditions of ribosome excess. In step C, an increase in NaCl (see text) decreases the pool of functional 40S ribosomal subunits, leaving the 5' noncoding sequence exposed. The consequent base pairing between complementary sequences, designated X and X', is illustrated in step D for constructs 1, 3, and 4 (see Fig. 2 for identification of these constructs). Whereas construct 1 cannot proceed beyond step D, with constructs 3 and 4 the subsequent advance of an 80S ribosome melts the base-paired region, allowing translation to continue. The key to the model is the ability of 80S elongating ribosomes to disrupt duplex structures that are too stable to be penetrated by 40S initiation complexes.

realization that 5'-3' complementarity can impair translation might explain why influenza viruses have evolved a way to avoid such structures in their mRNAs (9).

The recent literature contains several other examples of regulation by secondary structures in eucaryotic mRNAs. Translation of ferritin mRNA, for example, seems to be modulated by a weak hairpin structure near the 5' end of the message to which an iron-regulated repressor protein binds (27). A stem-loop structure at the 5' end of human immunodeficiency virus mRNA was reported to inhibit translation *in vitro* in *trans*, apparently by triggering the phosphorylation of eIF2 (7); *trans* inhibition did not occur with the hairpin structures used in the experiments described here, however. Very stable hairpin structures near the 5' ends of some other viral (32, 36) and cellular (26) mRNAs have been shown to inhibit translation by unexplored mechanisms. The highly conserved but not very stable secondary structure in the first exon of collagen mRNA would not be expected to, and did not, inhibit translation (1). In contrast with all of these negative effects of secondary structure, some recent experiments indicate that a stem-loop structure can enhance initiation when the hairpin occurs immediately downstream from an AUG codon around which the primary sequence is suboptimal for initiation (18). The realization that mRNA secondary structure can have positive as well as the more usual negative effects on translation, along with the fact that

the translatability of structure-prone mRNAs may vary with changes in the cytoplasmic environment (16; Fig. 2), underscores the complexity of the problem in eucaryotic cells. Notwithstanding such complications, the circumstances and mechanisms of translational regulation by mRNA secondary structure are slowly coming into view.

ACKNOWLEDGMENT

Research in my laboratory is supported by Public Health Service grant GM33915 from the National Institutes of Health.

LITERATURE CITED

1. Bornstein, P., J. McKay, S. Devarayalu, and S. C. Cook. 1988. A highly conserved, 5' untranslated, inverted repeat sequence is ineffective in translational control of the $\alpha 1(I)$ collagen gene. *Nucleic Acids Res.* **16**:9721-9736.
2. Bram, R. J., R. A. Young, and J. A. Steitz. 1980. The ribonuclease III site flanking 23S sequences in the 30S ribosomal precursor RNA of *E. coli*. *Cell* **19**:393-401.
3. Carter-Muenchau, P., and R. E. Wolf, Jr. 1989. Growth-rate dependent regulation of 6-phosphogluconate dehydrogenase level mediated by an anti-Shine-Dalgarno sequence located within the *Escherichia coli* *gnd* structural gene. *Proc. Natl. Acad. Sci. USA* **86**:1138-1142.
4. Christensen, A. K., L. E. Kahn, and C. M. Bourne. 1987. Circular polysomes predominate on the rough endoplasmic reticulum of somatotropes and mammatropes in the rat anterior pituitary. *Am. J. Anat.* **178**:1-10.
5. Cigan, A. M., L. Feng, and T. F. Donahue. 1988. tRNA^{met} functions in directing the scanning ribosome to the start site of translation. *Science* **242**:93-97.
6. Donahue, T. F., A. M. Cigan, E. K. Pabich, and B. C. Valavicius. 1988. Mutations at a Zn(II) finger motif in the yeast eIF-2 β gene alter ribosomal start-site selection during the scanning process. *Cell* **54**:621-632.
7. Edery, I., R. Petryshyn, and N. Sonenberg. 1989. Activation of double-stranded RNA-dependent kinase (dsI) by the TAR region of HIV-1 mRNA: a novel translational control mechanism. *Cell* **56**:303-312.
8. Hsu, M.-T., and M. Coca-Prados. 1979. Electron microscopic evidence for the circular form of RNA in the cytoplasm of eucaryotic cells. *Nature (London)* **280**:339-340.
9. Hsu, M.-T., J. D. Parvin, S. Gupta, M. Krystal, and P. Palese. 1987. Genomic RNAs of influenza viruses are held in a circular conformation in virions and in infected cells by a terminal panhandle. *Proc. Natl. Acad. Sci. USA* **84**:8140-8144.
10. Kozak, M. 1983. Translation of insulin-related polypeptides from messenger RNAs with tandemly reiterated copies of the ribosome binding site. *Cell* **34**:971-978.
11. 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.
12. Kozak, M. 1986. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eucaryotic ribosomes. *Cell* **44**:283-292.
13. Kozak, M. 1986. Influences of mRNA secondary structure on initiation by eucaryotic ribosomes. *Proc. Natl. Acad. Sci. USA* **83**:2850-2854.
14. Kozak, M. 1987. At least six nucleotides preceding the AUG initiator codon enhance translation in mammalian cells. *J. Mol. Biol.* **196**:947-950.
15. Kozak, M. 1987. An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res.* **15**:8125-8148.
16. Kozak, M. 1988. Leader length and secondary structure modulate mRNA function under conditions of stress. *Mol. Cell. Biol.* **8**:2737-2744.
17. Kozak, M. 1989. The scanning model for translation: an update. *J. Cell Biol.* **108**:229-241.
18. Kozak, M. 1989. Context effects and inefficient initiation at non-AUG codons in eucaryotic cell-free translation systems.

- Mol. Cell. Biol. 9:5073–5080.
19. Kozak, M., and A. J. Shatkin. 1976. Characterization of ribosome-protected fragments from reovirus messenger RNA. J. Biol. Chem. 251:4259–4266.
 20. Kozak, M., and A. J. Shatkin. 1977. Sequences of two 5'-terminal ribosome-protected fragments from reovirus messenger RNAs. J. Mol. Biol. 112:75–96.
 21. Kozak, M., and A. J. Shatkin. 1977. Sequences and properties of two ribosome binding sites from the small size class of reovirus messenger RNA. J. Biol. Chem. 252:6895–6908.
 22. Kozak, M., and A. J. Shatkin. 1978. Migration of 40S ribosomal subunits on messenger RNA in the presence of edeine. J. Biol. Chem. 253:6568–6577.
 23. Kruppa, J., and M. J. Clemens. 1984. Differential kinetics of changes in the state of phosphorylation of ribosomal protein S6 and in the rate of protein synthesis in MPC 11 cells during tonicity shifts. EMBO J. 3:95–100.
 24. Melton, D. A. 1985. Injected anti-sense RNAs specifically block messenger RNA translation in vivo. Proc. Natl. Acad. Sci. USA 82:144–148.
 25. Petersen, C. 1989. Long-range translational coupling in the *rplJL-rpoBC* operon of *Escherichia coli*. J. Mol. Biol. 206:323–332.
 26. Rao, C. D., M. Pech, K. C. Robbins, and S. A. Aaronson. 1988. The 5' untranslated sequence of the *c-sis*/platelet-derived growth factor 2 transcript is a potent translational inhibitor. Mol. Cell. Biol. 8:284–292.
 27. Rouault, T. A., M. W. Hentze, S. W. Caughman, J. B. Harford, and R. D. Klausner. 1988. Binding of a cytosolic protein to the iron-responsive element of human ferritin messenger RNA. Science 241:1207–1213.
 28. Saborio, J. L., S.-S. Pong, and G. Koch. 1974. Selective and reversible inhibition of initiation of protein synthesis in mammalian cells. J. Mol. Biol. 85:195–211.
 29. Shatkin, A. J. 1976. Capping of eucaryotic mRNAs. Cell 9:645–653.
 30. Shatkin, A. J. 1985. mRNA cap binding proteins: essential factors for initiating translation. Cell 40:223–224.
 31. Soares, M. B., E. Schon, A. Henderson, S. K. Karathanasis, R. Cate, S. Zeitlin, J. Chirgwin, and A. Efstratiadis. 1985. RNA-mediated gene duplication: the rat preproinsulin I gene is a functional retroposon. Mol. Cell. Biol. 5:2090–2103.
 32. Soe, L. H., C.-K. Shieh, S. C. Baker, M.-F. Chang, and M. M. C. Lai. 1987. Sequence and translation of the murine coronavirus 5'-end genomic RNA reveals the N-terminal structure of the putative RNA polymerase. J. Virol. 61:3968–3976.
 33. Spena, A., E. Krause, and B. Dobberstein. 1985. Translation efficiency of zein mRNA is reduced by hybrid formation between the 5'- and 3'-untranslated region. EMBO J. 4:2153–2158.
 34. Tinoco, I., P. N. Borer, B. Dengler, M. D. Levine, O. C. Uhlenbeck, D. M. Crothers, and J. Gralla. 1973. Improved estimation of secondary structure in ribonucleic acids. Nature (London) New Biol. 246:40–41.
 35. Walder, R. Y., and J. A. Walder. 1988. Role of RNase H in hybrid-arrested translation by antisense oligonucleotides. Proc. Natl. Acad. Sci. USA 85:5011–5015.
 36. Yager, D. R., and D. M. Coen. 1988. Analysis of the transcript of the herpes simplex virus DNA polymerase gene provides evidence that polymerase expression is inefficient at the level of translation. J. Virol. 62:2007–2015.