Minireview

Structural Features in Eukaryotic mRNAs That Modulate the Initiation of Translation*

Marilyn Kozak

From the Department of Biochemistry, University of Medicine and Dentistry of New Jersey, Piscataway, New Jersey 08854

In higher eukaryotes, translation is modulated at the level of initiation by five aspects of mRNA structure: (i) the m7G cap; (ii) the primary sequence or context surrounding the AUG codon; (iii) the position of the AUG codon, i.e., whether or not it is "first"; (iv) secondary structure both upstream and downstream from the AUG codon; and (v) leader length. Here I briefly discuss how experimental manipulation of these features affects the fidelity and/or efficiency of initiation. Elsewhere (1) I discuss the extent to which natural mRNA leader sequences conform to these experimentally determined requirements for initiation.

Although my primary concern is to document the occurrence and consequences of the five structural elements in mRNAs, I will allude now and then to why each feature has the effect it does. The explanations will invariably hark back to the scanning process by which ribosomes are thought to initiate translation. In its simplest form, the scanning model (2) postulates that a 40 S ribosomal subunit, carrying Met-tRNA\(^{\text{Met}}\) and an imperfectly defined set of initiation factors (3), enters at the 5'-end of the mRNA and migrates linearly until it reaches the first AUG codon, whereupon a 60 S ribosomal subunit joins and the first peptide bond is formed (Fig. 1). Evidence in support of the scanning model has been added previously (2, 4, 5). The most recent evidence includes the apparent queuing of 40 S ribosomal subunits on a long, unstructured leader sequence (6) and the demonstration that 40 S subunits stall on the 5'-side of a stable hairpin structure introduced between the cap and the AUG codon (7).

m7G Cap

That the ubiquitous m7G cap increases the efficiency of translation in vitro was first shown by Shatkin (8) and has been confirmed many times since. While the dependence of in vitro translation on the m7G cap may vary with the choice of reaction conditions, in vivo translation of most mRNAs is stringently cap-dependent, as shown in studies with vesicular stomatitis virus mutants that are defective in methylation (9) and with various other test systems (10, 11). Parenthetically, the experiments with vesicular stomatitis virus distinguish nicely between the ability of the cap to stabilize transcripts and its ability to stimulate translation; a guanylylated, unmethylated cap is sufficient to protect transcripts from 5'-exonucleases, but N-7 methylation is essential for efficient translation. A methylated cap and the associated cap-binding protein may be less important for the translation of mRNAs that have a rather long, unstructured leader sequence (6, 12), but such mRNAs are rare. Among animal cells and viruses, the only mRNAs that clearly are translated without benefit of a cap derive from picornaviruses.

Context

Systematic mutagenesis of nucleotides in the vicinity of the AUG codon revealed that GCC\(^{\text{GCU}}\)UGG is the optimal context for initiation of translation in cultured monkey cells (13, 14). (The A of the AUG codon is designated +1, with positive and negative integers proceeding 3' and 5', respectively.) A purine, preferably A, in position -3 and a G in position +4 have the strongest effects, modulating translation at least 10-fold; the smaller effects of other nucleotides near the AUG codon are seen most easily in the absence of A-3 and G+4. To be effective, the GCCACC motif must abut the AUG codon. Shifting the motif by just one nucleotide to the left or right abolishes its facilitating effect (13, 14). The strong contributions of A or G in position -3 and G in position +4, deduced initially in transfection assays with COS cells, have been confirmed in experiments with transformed plants (15, 16) and with standard in vitro translation systems from plants (17) and animals (17-21). One set of constructs used for the in vitro translation experiments was designed with two in-frame AUG codons, positioned to produce "long" and "short" versions of chloramphenicol acetyltransferase (17). The experiments carried out with those constructs revealed that a suboptimal context around the first AUG codon causes some 40 S ribosomal subunits to bypass the first AUG and initiate instead at the second AUG codon. Thus, context affects the fidelity as well as the efficiency of initiation. When the first AUG codon lies in a weak context, it is recognized inefficiently irrespective of the mRNA concentration and irrespective of the presence or absence of competing mRNAs (17), consistent with the scanning model which postulates that recognition of the AUG codon occurs after the competition-sensitive binding of the 40 S ribosome to the 5'-end of the mRNA. The leaky scanning that results from a suboptimal context around the first AUG codon enables some viral mRNAs to produce two proteins by initiating at the first and second AUG codons, as described elsewhere (1, 22).

The experimentally determined optimal context for initiation (13, 14) matches the consensus sequence derived from inspection of 699 vertebrate mRNA sequences (23). Except for yeasts (24), other eukaryotic organisms that have been examined show context effects similar to those described for vertebrates. Thus, ACC in positions -3 to -1 promotes translation of chloramphenicol acetyltransferase in Drosophila (25), although the actual consensus sequence in flies differs slightly from mammals. Plant mRNAs have the expected purine in position -3 (in 93% of 252 mRNA sequences examined) and the expected G in position +4 (in 74% of the mRNAs examined); and those conserved nucleotides augment translation in plants (15-17). A recent compilation of translational start sites in protozoa also shows a strong preference for A in position -3 (26).

Position

The scanning model predicts that ribosomes should initiate at the first AUG codon in a good context, a prediction that is upheld by most (perhaps 90%) of vertebrate mRNAs (23). (The number cannot be fixed more precisely because reports of CDNA sequences with upstream AUG codons often turn out to reflect errors or misinterpretations, as documented elsewhere (1, 2, 23).) The importance of position in determin-

* Research in my laboratory is supported by National Institutes of Health Grant GM33915.

1 M. Kozak, unpublished compilation.
Surprising to find that a small amount of secondary structure near the start of the coding sequence can actually enhance recognition of the preceding AUG codon. Downstream secondary structure apparently contributes to the fidelity of initiation by preventing the 40 S ribosome from scanning too fast or too far. The effect is striking in that the introduction of downstream secondary structure (ΔG, −19 kcal/mol) can completely suppress the leaky scanning that otherwise occurs when the first AUG codon lies in an unfavorable context (36). The simplest rationalization is that downstream secondary structure slows scanning, thereby providing more time for recognition of the preceding AUG codon. The maximal effect occurs when 14 nucleotides intervened between the base of the hairpin and the preceding AUG codon, and that minimum fits nicely with RNase protection experiments which mapped the leading edge of an initiating ribosome 12–15 nucleotides 3′ of the AUG codon. Thus, our working hypothesis is that a hairpin located 12–15 nucleotides downstream from the AUG codon causes the 40 S ribosomal subunit to stall momentarily with its AUG-recognition center right over the AUG codon, thereby facilitating initiation.

There are several situations in which this positive effect of secondary structure might be important. Although 97% of mammal mRNAs that require the AUG codon for initiation have very few possess the full consensus sequence. Thus, some feature in addition to primary structure would seem to be required to explain the usual absence of leakiness, and the possibility that secondary structure near the start of the coding sequence compensates for the less-than-perfect context around the AUG codon seems an attractive solution. The contribution of downstream secondary structure might be especially important for the handful of vertebrate mRNAs that initiate translation at an AUG codon in a very weak context (i.e. lacking both a purine in position −3 and G in position +4 (23)) or at a non-AUG codon. The experimental imposition of appropriately positioned secondary structure indeed increases initiation from cryptic non-AUG codons in test cases (36), and nearly all of the mRNAs that naturally support initiation from upstream non-AUG codons have extraordinarily GC-rich (hence highly structured) leader sequences (Fig. 2).

**Secondary Structure: Negative Effects**

In contrast with the positive effects of secondary structure introduced downstream from the AUG codon, stem-and-loop structures introduced between the cap and the AUG codon generally impair translation. Whether or not the structure upstream from the AUG codon impairs translation depends on the strength and position of the hairpin. A summary of the rules follows. (i) A modest amount of secondary structure near the cap (i.e. within the first 12 nucleotides) can drastically inhibit translation (7). Secondary structure in this position has been shown to prevent mRNA from binding to 40 S ribosomes (7), as expected if the 5′-end of the mRNA is the entry site for ribosomes (Fig. 1). The much discussed inhibition of ferritin mRNA translation by the IRE-binding protein fits nicely with RNase protection experiments which mapped the leading edge of a hairpin 12–15 nucleotides 3′ of the AUG codon. Thus, our working hypothesis is that a hairpin located 12–15 nucleotides downstream from the AUG codon causes the 40 S ribosomal subunit to stall momentarily with its AUG-recognition center right over the AUG codon, thereby facilitating initiation.

There are several situations in which this positive effect of secondary structure might be important. Although 97% of vertebrate mRNAs that require the AUG codon for initiation have very few possess the full consensus sequence. Thus, some feature in addition to primary structure would seem to be required to explain the usual absence of leakiness, and the possibility that secondary structure near the start of the coding sequence compensates for the less-than-perfect context around the AUG codon seems an attractive solution. The contribution of downstream secondary structure might be especially important for the handful of vertebrate mRNAs that initiate translation at an AUG codon in a very weak context (i.e. lacking both a purine in position −3 and G in position +4 (23)) or at a non-AUG codon. The experimental imposition of appropriately positioned secondary structure indeed increases initiation from cryptic non-AUG codons in test cases (36), and nearly all of the mRNAs that naturally support initiation from upstream non-AUG codons have extraordinarily GC-rich (hence highly structured) leader sequences (Fig. 2).

**Secondary Structure: Positive Effects**

While there have long been hints that secondary structure in mRNAs can reduce the efficiency of translation, it was surprising to find that a small amount of secondary structure near the start of the coding sequence can actually enhance recognition of the preceding AUG codon. Downstream secondary structure apparently contributes to the fidelity of initiation by preventing the 40 S ribosome from scanning too fast or too far. The effect is striking in that the introduction of downstream secondary structure (ΔG, −19 kcal/mol) can completely suppress the leaky scanning that otherwise occurs when the first AUG codon lies in an unfavorable context (36). The simplest rationalization is that downstream secondary structure slows scanning, thereby providing more time for recognition of the preceding AUG codon. The maximal effect occurs when 14 nucleotides intervened between the base of the hairpin and the preceding AUG codon, and that minimum fits nicely with RNase protection experiments which mapped the leading edge of an initiating ribosome 12–15 nucleotides 3′ of the AUG codon. Thus, our working hypothesis is that a hairpin located 12–15 nucleotides downstream from the AUG codon causes the 40 S ribosomal subunit to stall momentarily with its AUG-recognition center right over the AUG codon, thereby facilitating initiation.

There are several situations in which this positive effect of secondary structure might be important. Although 97% of vertebrate mRNAs that require the AUG codon for initiation have very few possess the full consensus sequence. Thus, some feature in addition to primary structure would seem to be required to explain the usual absence of leakiness, and the possibility that secondary structure near the start of the coding sequence compensates for the less-than-perfect context around the AUG codon seems an attractive solution. The contribution of downstream secondary structure might be especially important for the handful of vertebrate mRNAs that initiate translation at an AUG codon in a very weak context (i.e. lacking both a purine in position −3 and G in position +4 (23)) or at a non-AUG codon. The experimental imposition of appropriately positioned secondary structure indeed increases initiation from cryptic non-AUG codons in test cases (36), and nearly all of the mRNAs that naturally support initiation from upstream non-AUG codons have extraordinarily GC-rich (hence highly structured) leader sequences (Fig. 2).

**Secondary Structure: Negative Effects**

In contrast with the positive effects of secondary structure introduced downstream from the AUG codon, stem-and-loop structures introduced between the cap and the AUG codon generally impair translation. Whether or not the structure upstream from the AUG codon impairs translation depends on the strength and position of the hairpin. A summary of the rules follows. (i) A modest amount of secondary structure near the cap (i.e. within the first 12 nucleotides) can drastically inhibit translation (7). Secondary structure in this position has been shown to prevent mRNA from binding to 40 S ribosomes (7), as expected if the 5′-end of the mRNA is the entry site for ribosomes (Fig. 1). The much discussed inhibition of ferritin mRNA translation by the IRE-binding protein fits nicely with RNase protection experiments which mapped the leading edge of a hairpin 12–15 nucleotides 3′ of the AUG codon. Thus, our working hypothesis is that a hairpin located 12–15 nucleotides downstream from the AUG codon causes the 40 S ribosomal subunit to stall momentarily with its AUG-recognition center right over the AUG codon, thereby facilitating initiation.

There are several situations in which this positive effect of secondary structure might be important. Although 97% of vertebrate mRNAs that require the AUG codon for initiation have very few possess the full consensus sequence. Thus, some feature in addition to primary structure would seem to be required to explain the usual absence of leakiness, and the possibility that secondary structure near the start of the coding sequence compensates for the less-than-perfect context around the AUG codon seems an attractive solution. The contribution of downstream secondary structure might be especially important for the handful of vertebrate mRNAs that initiate translation at an AUG codon in a very weak context (i.e. lacking both a purine in position −3 and G in position +4 (23)) or at a non-AUG codon. The experimental imposition of appropriately positioned secondary structure indeed increases initiation from cryptic non-AUG codons in test cases (36), and nearly all of the mRNAs that naturally support initiation from upstream non-AUG codons have extraordinarily GC-rich (hence highly structured) leader sequences (Fig. 2).

**Secondary Structure: Positive Effects**

While there have long been hints that secondary structure in mRNAs can reduce the efficiency of translation, it was surprising to find that a small amount of secondary structure near the start of the coding sequence can actually enhance recognition of the preceding AUG codon. Downstream secondary structure apparently contributes to the fidelity of initiation by preventing the 40 S ribosome from scanning too fast or too far. The effect is striking in that the introduction of downstream secondary structure (ΔG, −19 kcal/mol) can completely suppress the leaky scanning that otherwise occurs when the first AUG codon lies in an unfavorable context (36). The simplest rationalization is that downstream secondary structure slows scanning, thereby providing more time for recognition of the preceding AUG codon. The maximal effect occurs when 14 nucleotides intervened between the base of the hairpin and the preceding AUG codon, and that minimum fits nicely with RNase protection experiments which mapped the leading edge of an initiating ribosome 12–15 nucleotides 3′ of the AUG codon. Thus, our working hypothesis is that a hairpin located 12–15 nucleotides downstream from the AUG codon causes the 40 S ribosomal subunit to stall momentarily with its AUG-recognition center right over the AUG codon, thereby facilitating initiation.

There are several situations in which this positive effect of secondary structure might be important. Although 97% of vertebrate mRNAs that require the AUG codon for initiation have very few possess the full consensus sequence. Thus, some feature in addition to primary structure would seem to be required to explain the usual absence of leakiness, and the possibility that secondary structure near the start of the coding sequence compensates for the less-than-perfect context around the AUG codon seems an attractive solution. The contribution of downstream secondary structure might be especially important for the handful of vertebrate mRNAs that initiate translation at an AUG codon in a very weak context (i.e. lacking both a purine in position −3 and G in position +4 (23)) or at a non-AUG codon. The experimental imposition of appropriately positioned secondary structure indeed increases initiation from cryptic non-AUG codons in test cases (36), and nearly all of the mRNAs that naturally support initiation from upstream non-AUG codons have extraordinarily GC-rich (hence highly structured) leader sequences (Fig. 2).
The recent report (55) that structure-prone mRNAs are translated with, as in baculovirus and riboprobe vectors, the introduction cell types they also switched vectors. (If the vector-derived inhibition more profoundly than if a bit more secondary structure might be expected to indicate some novel, developmentally regulated helicase activity, although it is possible that covalent modification of hairpin-containing mRNAs by a previously recognized unwinding activity (56) is responsible for the enhanced translation.

**Minireview: mRNA Structure and the Initiation of Translation**

**Leader Length**

Recognition of the first AUG codon may be impaired when it is positioned too close to the cap (57-59). When this issue was explored systematically, using synthetic transcripts in which the first AUG codon was in a favorable context, about half of the ribosomes bypassed the first AUG codon and initiated instead at a downstream site when the first AUG occurred within 12 nucleotides of the cap (60). The leakiness was suppressed when the leader sequence was lengthened to 20 nucleotides or when a modest amount of secondary structure was introduced downstream from the first AUG codon (60). In natural mRNAs the amount of secondary structure near the beginning of the coding region varies, making it hard to say a priori whether a particular short leader sequence will pose a problem.

Further lengthening of the 5'-noncoding sequence beyond the 20 or so nucleotides required for the fidelity of initiation can dramatically increase the efficiency of translation in vitro (6). The increased efficiency was clearly attributable to leader length, rather than to any particular sequence, inasmuch as insertion of three different synthetic oligonucleotides, each 60 nucleotides long, stimulated translation identically (6). The only feature common to all three sequences was a paucity of nucleotides, which ensured against the formation of secondary structure. The efficiency of translation in vitro was proportional to leader length in the range of 17 to about 80 nucleotides (6). Augmentation of translation by long, synthetic leader sequences mimics the effects of certain natural leader sequences (61-65). The fact that the precise sequence of these translational "enhancers" is not critical (6, 61-64) makes it unlikely that their facilitating effect on translation is mediated by proteins that recognize particular sequence motifs. Rather, the observed loading of extra 40 S ribosomal subunits on long leader sequences (6) seems likely to underlie the improvement in translation. In keeping with that interpretation, a long unstructured leader sequence augments translation only when it is at the exact 5'-end of the transcript (6, 53, 63). The introduction of a moderately long, unstructured synthetic leader sequence turns out to be an easy way to increase the efficiency of in vitro expression vectors (6).
**Minireview: mRNA Structure and the Initiation of Translation**

### Closing Notes

The trick to identifying elements within 5'-noncoding sequences that modulate translation is to isolate each feature. There are many ways that the effects of context, upstream AUG codons, etc., might not be seen. Mutations that change the primary sequence around the initiator codon may have little effect if the sequence 3' of the AUG codon is structured, inasmuch as downstream secondary structure compensates for absence of the preferred context (36). Since features in addition to context modulate initiation, as described herein, it follows that virtually nothing can be learned from comparing two completely different leader sequences in which context is only one of many variables (66). When matched mRNAs are compared, converting a good context to a poorer one usually reduces the translational yield, but converting a poor context to a better one may not increase the yield of protein if some other step (such as elongation (67) or protein processing (68)) is limiting. Biological assays, albeit very sensitive, are so many steps removed from translation that failure to see the expected effects of context are difficult to interpret (69).

In cases where less mRNA accumulates in cells under conditions (such as the imposition of a poor context) that impair translation, the gesture of "correcting" protein yields for differences in mRNA levels may ablate the perceived effects of context on translation (25). Thus, the best systems for testing effects of context are those in which differences in translatability do not affect mRNA stability. Context effects if inappropriate reaction conditions are used for testing effects of context are those in which differences in context to a better one may not increase the yield of protein greater than it really is.

The extent to which natural mRNA leader sequences contain elements that support the proper selection of initiator codons is only one of many variables (66). When matched mRNAs that are 5'-translated by two completely different leader sequences in which context is a variable. Because the consensus sequence for initiation is GC-rich, mutations that improve the primary structure may inadvertently increase secondary structure to a point that becomes inhibitory. Some evidence (26, 27) is only one of many variables (66).

There are many ways that the effects of context, upstream AUG codons, etc., might not be seen. Mutations that change the primary sequence around the initiator codon may have little effect if the sequence 3' of the AUG codon is structured, inasmuch as downstream secondary structure compensates for absence of the preferred context (36). Since features in addition to context modulate initiation, as described herein, it follows that virtually nothing can be learned from comparing two completely different leader sequences in which context is only one of many variables (66).

### References

43. Interpretation can be complicated even when context is the only variable. Because the consensus sequence for initiation is GC-rich, mutations that improve the primary structure may inactivate secondary structure to a point that becomes inhibitory. Some evidence of this was seen in Ref. 41 and the facilitating effect of context could nevertheless be discerned in those experiments by monitoring initiation from the first versus the second AUG codon.
44. The problem is that a defect in translation sometimes accelerates mRNA degradation, in which case expressing the yield of protein as a function of the (lowered) steady-state level of mRNA makes the efficiency of translation appear greater than it really is.