MOLECULAR AND CELLULAR BIOLOGY, Nov. 1989, p. 5073–5080
0270-7306/89/115073-08$02.00/0

Vol. 9, No. 11

Context Effects and Inefficient Initiation at Non-AUG Codons in Eucaryotic Cell-Free Translation Systems

MARILYN KOZAK

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

Received 5 June 1989/Accepted 1 August 1989

The context requirements for recognition of an initiator codon were evaluated in vitro by monitoring the relative use of two AUG codons that were strategically positioned to produce long (pre-chloramphenicol acetyl transferase [CAT]) and short versions of CAT protein. The yield of pre-CAT initiated from the 5′-proximal AUG codon increased, and synthesis of CAT from the second AUG codon decreased, as sequences flanking the first AUG codon increasingly resembled the eucaryotic consensus sequence. Thus, under prescribed conditions, the fidelity of initiation in extracts from animal as well as plant cells closely mimics what has been observed in vivo. Unexpectedly, recognition of an AUG codon in a suboptimal context was higher when the adjacent downstream sequence was capable of assuming a hairpin structure than when the downstream region was unstructured. This finding adds a new, positive dimension to regulation by mRNA secondary structure, which has been recognized previously as a negative regulator of initiation. Translation of pre-CAT from an AUG codon in a weak context was not preferentially inhibited under conditions of mRNA competition. That result is consistent with the scanning model, which predicts that recognition of the AUG codon is a late event that occurs after the competition-sensitive binding of a 40S ribosome-factor complex to the 5′ end of mRNA. Initiation at non-AUG codons was evaluated in vitro and in vivo by introducing appropriate mutations in the CAT and preproinsulin genes. GUG was the most efficient of the six alternative initiator codons tested, but GUG in the optimal context for initiation functioned only 3 to 5% as efficiently as AUG. Initiation at non-AUG codons was artifactually enhanced in vitro at supraoptimal concentrations of magnesium.

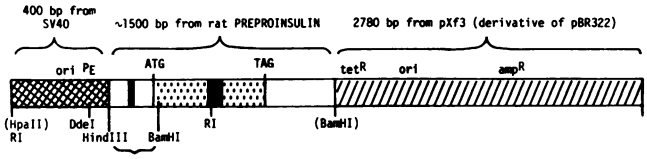
In the absence of objective criteria for deciding the correct reaction conditions for translation in vitro, almost any answer is possible to the question of whether a particular AUG or non-AUG codon can initiate protein synthesis. The common practice of adjusting the Mg^{2+} concentration for optimal incorporation of amino acids seems inappropriate for studies on the mechanism of initiation, since the reaction conditions that support optimal amino acid incorporation are usually a compromise between the requirements for initiation and the quite different requirements for elongation (57). Even when the objective is purely practical—using a cell-free translation system to detect the polypeptide encoded by a cDNA clone—production of the desired polypeptide may be impaired if the reaction conditions allow extraneous upstream or downstream initiation events. Here I propose two criteria for standardizing translation extracts: namely, that in vitro translation systems should reflect the same context requirements (28, 31) and the same low efficiency of initiation at non-AUG codons as occurs in vivo. I show that when the Mg^{2+} and spermidine concentrations are appropriately adjusted, both the popular wheat germ and rabbit reticulocyte translation systems can be made to meet those requirements.

Using the reaction conditions thus defined, I have studied the relationship between AUG codon context and mRNA competition. I have also examined the possibility that secondary structure near the beginning of the coding sequence might, by slowing the advance of the 40S ribosomal subunit, enhance the recognition of an AUG codon around which the primary sequence is suboptimal for initiation. The results of these experiments are explicable in light of the scanning model for initiation in eucaryotes (33).

MATERIALS AND METHODS

Construction and purification of plasmids. For in vivo expression experiments, derivatives of a simian virus 40-based preproinsulin plasmid were obtained by substituting various synthetic oligonucleotides for the 178-base-pair, AUG-containing *Hin*dIII-to-*Bam*HI fragment of p255/11 (Fig. 1). Complementary deoxyoligonucleotides, of which only the sense strands are shown in Fig. 1, were annealed and then ligated into the *Hin*dIII-*Bam*HI-linearized acceptor as described previously (28). For in vitro expression experiments, the starting construct was a derivative of the Riboprobe vector pSP64 (Promega Corp.) into which a chloramphenicol acetyltransferase (CAT) cartridge (Pharmacia, Inc.) was inserted at the *Bam*HI site. Constructs in which an upstream, in-frame AUG (or other potential initiator) codon supplanted the normal initiator codon for CAT were obtained by replacing the *Hin*dIII-to-*Bam*HI portion of the pSP64 polylinker (shown in lowercase letters in Fig. 2A) with an appropriate synthetic deoxyoligonucleotide (Fig. 2B).

Transformation of competent *Escherichia coli* MM294 cells was carried out as described previously (26). Plasmids were amplified by adding either chloramphenicol (170 μg/ml, for preproinsulin constructs) or spectinomycin (300 μg/ml, for SP64 · CAT derivatives) to log-phase cultures in M9 medium. Highly purified plasmid DNA for COS transfection assays and SP6 transcription reactions was prepared by lysozyme-detergent lysis of bacteria, followed by phenol extraction, Bio-Gel A50 (Bio-Rad Laboratories) chromatography, and banding in CsCl gradients in the presence of ethidium bromide. The structures of new plasmids were confirmed by dideoxy-chain termination sequencing reac-



G1	agctccaccTTGgg	B13	agcttccaccATGgccctgtg
G2	agctccaccATTgg	B61	agcttccaccGTGgccctgtg
G3	agctccaccATAgg	B62	agcttgggttGTGgccctgtg
G4	agctccaccATGgg	B64	agcttggtttGTGgccctgtg
G5	agctccaccCTGgg	G6	agctccaccACGgg

FIG. 1. Construction of mutants used to evaluate initiation at non-AUG codons in vivo. The parental plasmid p255/11, shown on the top line, has been described previously (27). Cleavage of p255/11 at the unique *Hin*dIII and *Bam*HI sites releases a small fragment that contains the initiator codon for preproinsulin. In its place, one of the indicated synthetic oligonucleotides was inserted. Sequences are depicted with T in place of U here and in Fig. 2.

tions performed with Sequenase-2 (U.S. Biochemical Corp.) (67). In Fig. 1 and 2, nucleotide sequences are shown with T in place of U because it was DNA that I actually manipulated and sequenced. In the other figures and text, it seemed to make more sense to cite the corresponding RNA sequences.

Transfection of COS cells for in vivo translation studies. The transfection technique has been described (26, 28), 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 of SP64 · CAT constructs. Transcription of *Ava*I-linearized plasmid DNA with SP6 polymerase (Pharmacia) was done by the general procedures of Melton et al. (42). Capped transcripts were obtained by supplementing the cocktail with m^7GpppG (500 μM; Pharmacia) and lowering the GTP concentration to 20 μM; ATP, CTP, and UTP were used at 500 μM. To ensure uniformity, aliquots of a common reaction mixture were used to synthesize all of the transcripts intended for a particular translation experiment. The transcription reactions included [^{3}H]UTP (48 Ci/mmol; 20 μCi/100-μl reaction mixture), which made it possible to measure exactly the amount of mRNA used for each translation assay. Transcription reactions were terminated by digesting the templates with RQ1 DNase. The mRNAs were extracted with phenol before application to prespun Sephadex G-50 columns (Boehringer Mannheim Biochemicals).

Each 25-μl translation reaction mixture contained 40 μCi of [^{35}S]methionine (1,000 Ci/mmol; Dupont, NEN Research Products) and 0.1 to 0.2 μg of mRNA. Incubation was for 60 min at 20°C (wheat germ) or 30°C (reticulocyte extracts). The reticulocyte translation kit from Bethesda Research Laboratories, Inc, (8125SA) was supplemented to achieve final concentrations of 2.2 mM Mg^{2+}, 45 mM KCl, and 90 mM potassium acetate. With wheat germ extracts, which were prepared in house, the 25-μl reaction mixture contained 10 μl of S30 extract, 30 mM *N*-2-hydroxyethylpiperazine-*N'*-2-ethanesulfonic acid (HEPES; pH 7.5), 72 mM KCl, 8 mM creatine phosphate, 3 μg of creatine phosphokinase, 0.2 mM GTP, 1 mM ATP, 2 mM dithiothreitol, and 19 unlabeled amino acids at 50 μM each. The concentrations of Mg^{2+} and spermidine and the source of mRNA are indicated in the

A. 5' noncoding sequences of parental SP64 vectors:

```
                  Start SP6 transcription                                          Start CAT
                  ↓     HindIII          BamHI                                     translation
                  ↓     ↓                ↓                                         ↓
SP64·CAT          GAATACAagcttgggctgcaggtcgactctagag----------GATCCGAGATTTTCAGGAGCTAAGGAAGCTAAAATGG

SP64(8334)CAT     GAATACAagcttgggctgcaggtcgactctagagGATCCAAAGACTGCCAAATCTAGATCCGAGATTTTCAGGAGCTAAGGAAGCTAAAATGG

SP64(8336)CAT     GAATACAagcttgggctgcaggtcgactctagag←GATCCGGGTTCTCCCGGATC→AAGATCCGAGATTTTCAGGAGCTAAGGAAGCTAAAATGG
```

B. Oligonucleotides substituted for the HindIII-to-BamHI portion of SP64(8334)CAT or SP64(8336)CAT:

B38 - agcttggtttATGtccctgtg	B171 - agcttgccgccgccATGtcagg	B61 - agcttccaccGTGgccctgtg
B34 - agcttggtttATGgccctgtg	B169 - agctttcctcctccATGtcagg	B64 - agcttggtttGTGgccctgtg
B13 - agcttccaccATGgccctgtg	B173 - agcttgccgccacccATGtcagg	B70 - agcttccaccTTGgccctgtg
		B80 - agcttccaccCTGgccctgtg

C. 5' sequence of SP64(8336)B34 and related constructs:

```
SP64(8336)B34     GAATACAagcttggtttATGgccctgtg←GATCCGGGTTCTCCCGGATC→AA/GATCCGAGATTTTCAGGAGCTAAGGAAGCTAAAATGG...

SP64(8414)B34     GAATACAagcttggtttATGgccctgtg←GATCCCCCTTCTCCC→GGATCAA/GATCCGAGATTTTCAGGAGCTAAGGAAGCTAAAATGG...

SP64(8416)B34     GAATACAagcttggtttATGgccctgtg←GATCCCCCTTCTGGGGGATC→AA/GATCCGAGATTTTCAGGAGCTAAGGAAGCTAAAATGG...

SP64(8334)B34     GAATACAagcttggtttATGgccctgtgGATCCAAAGACTGCCAAATCTA/GATCCGAGATTTTCAGGAGCTAAGGAAGCTAAAATGG...
                  ↑                ↑                                                ↑
                  Start SP6        Start pre-CAT                                    Start CAT
                  transcription    translation                                      translation
```

FIG. 2. Construction of SP6 derivatives for in vitro transcription and translation. (A) 5' noncoding sequence of the parental construct SP64 · CAT (line 1). Two intermediate constructs, SP64(8334)CAT and SP64(8336)CAT, were derived from SP64 · CAT by inserting a 22-nucleotide adaptor (lines 2 and 3). (B) Oligonucleotides that replaced the *Hin*dIII-to-*Bam*HI portion of SP64(8334 or 8336)CAT; the AUG codon carried by these inserts initiated the N-terminally extended pre-CAT protein. (C) Full sequence at the 5' end of SP64(8336)B34 along with sequences of the matched constructs used in Fig. 6.

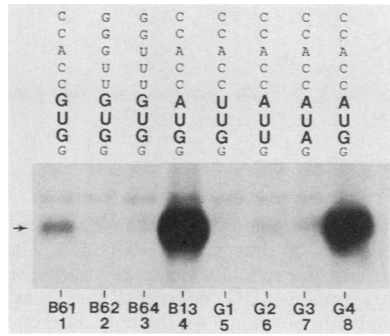


FIG. 3. Transfection of COS cells with plasmids in which GUG, UUG, AUU, or AUA replaced the AUG initiator codon for preproinsulin. The yield of [^{35}S]cysteine-labeled proinsulin from the AUG-containing control plasmids (lanes 4 and 8) was 20 times that obtained with B61 (lane 1). A cropped fluorogram is shown of a polyacrylamide gel that had been exposed for 3 weeks. →, Position of proinsulin.

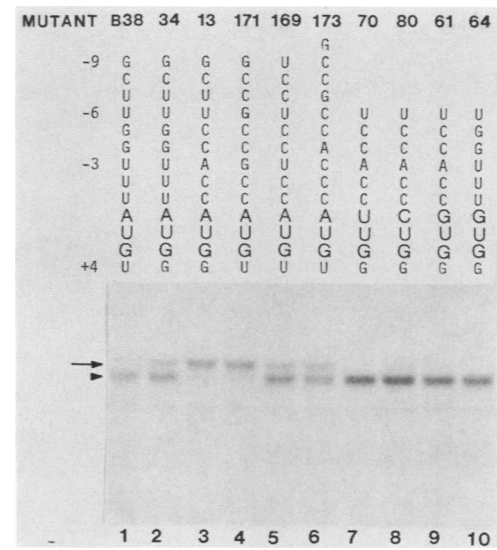


FIG. 4. Context-dependent recognition of AUG and non-AUG initiator codons by reticulocyte ribosomes. Equal amounts of SP6 vector-derived transcripts were translated in a reticulocyte lysate at 2.2 mM Mg^{2+}. An autoradiogram of [^{35}S]methionine-labeled proteins, fractionated by polyacrylamide gel electrophoresis, is shown. The sequence around the 5′-proximal initiator codon, which is the pre-CAT start site, is shown above each lane. All mRNAs contained oligonucleotide 8334 as the adaptor. Here and in all subsequent figures, an arrow at the left marks the position of pre-CAT initiated from the upstream start site; an arrowhead marks the position of wild-type CAT protein.

appropriate figure legends. At the end of the incubation, 5 μl of each reaction mixture was mixed with 5 μl of electrophoresis buffer (36) supplemented with 20% glycerol, 10% β-mercaptoethanol, and 0.01% bromophenol blue. The samples were heated for 7 min at 95°C and loaded onto a 15% polyacrylamide–sodium dodecyl sulfate–6 M urea gel. Autoradiograms of dried gels were developed after exposure for about 4 days. Protein bands were quantitated by densitometry.

RESULTS

Ability of non-AUG codons to initiate translation in mammalian cells. As a prelude to studying initiation at non-AUG codons in vitro, I evaluated the ability of other codons to initiate the translation of preproinsulin in transfected COS cells. Of the six non-AUG codons that were tested (GUG, UUG, AUU, AUA, CUG, and ACG; Fig. 1), GUG functioned most efficiently in vivo; but GUG, in the most favorable context for initiation, was only 3 to 5% as active as AUG in the same context (B61 versus B13 in Fig. 3). The conclusion that GUG is the functional, albeit weak, initiator codon in plasmid B61 seems justified by the failure of B62 and B64 to produce any proinsulin (Fig. 3, lanes 2 and 3): the only difference among the three plasmids is the block of 5 nucleotides immediately upstream from the GUG codon. Since GUG initiated translation, albeit inefficiently, only when it was preceded by CCACC, initiation at a GUG codon in vivo shows the same dependence on context as does initiation at AUG codons (28). None of the other constructs listed in Fig. 1 was more active than B61, although a faint band of proinsulin was detectable with G5 (CUG) and G6 (ACG).

In reticulocyte lysates, recognition of initiator codons depends on the flanking primary sequence. SP6 vector-derived transcripts were used to determine whether cell-free translation systems would show the same context requirements and the same low ability to initiate at non-AUG codons as whole cells. The control construct SP64 · CAT, which has no upstream AUG triplets, initiates translation from the authentic CAT start site (marked "Start CAT" in Fig. 2A). The other constructs have an upstream in-frame AUG (or alternative) initiator codon that was introduced on one of the inserts depicted in Fig. 2B; an adaptor, oligonucleotide 8334 or 8336, was used to move the upstream AUG codon into the same reading frame as CAT. To the extent that the upstream AUG (or alternative) codon can support initiation, the derivatives listed in Fig. 2B should produce a CAT-related polypeptide with an N-terminal extension of 22 amino acids. The data in Fig. 4 to 7 show that the elongated protein, designated pre-CAT, was adequately resolved from wild-type CAT.

Figure 4 shows the results of an experiment in which derivatives of SP64(8334)CAT were translated in a messenger-dependent reticulocyte lysate. The net Mg^{2+} and potassium concentrations were 2.2 and 135 mM, respectively. In vitro synthesis of pre-CAT increased at least sixfold (Fig. 4, lane 1 versus lanes 3 and 4), and the yield of CAT initiated from the second AUG codon simultaneously decreased, as the nucleotides flanking the 5′-proximal AUG codon were brought into conformity with the consensus sequence for initiation. The threefold-higher yield of pre-CAT obtained with SP64(8334)B34 than with SP64(8334)B38 indicates the important contribution of G in position +4. The sixfold higher yield of pre-CAT from SP64(8334)B171 than from SP64(8334)B38 indicates that even in the absence of G in position +4, initiation is very efficient when GCCGCCGCC precedes the AUG codon provided that the purines fall in positions −3, −6, and −9. In contrast, a construct in which the purines were shifted to positions −4, −7, and −10 [SP64(8334)B173] was as leaky as a construct that had no upstream purines [SP64(8334)B169]. Initiation at upstream non-AUG codons in these constructs was appropriately inefficient as evidenced by the barely detectable pre-CAT protein in lanes 7 to 10 of Fig. 4. The reproducibility of results from the in vitro translation assay is illustrated in Fig. 5A (lanes 1 to 8), which shows duplicate reactions with each mRNA (adjacent lanes, bracketed).

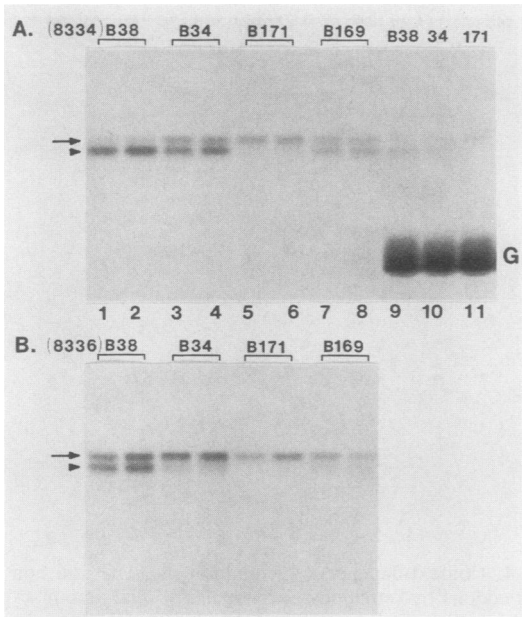


FIG. 5. Demonstration that sequences 3′ to the AUG codon modulate recognition of the codon by reticulocyte ribosomes. The experiment was carried out as for Fig. 4, using mRNAs from series 8334 (A) or 8336 (B). Each mRNA was tested in duplicate (adjacent lanes, bracketed) to illustrate the reproducibility of the assay. Rabbit globin mRNA (0.5 μg) was included as competitor in the lanes 9 to 11 of panel A. G, Globin translation product.

Downstream secondary structure enhances recognition of the preceding AUG codon. Whereas the foregoing results were predictable from prior in vivo demonstrations of context effects on initiation (28, 31), the results shown in Fig. 5B with a parallel set of constructs revealed something unexpected: sequences downstream from an AUG codon could also affect recognition of the codon. Thus, when the adaptor following the first AUG codon was switched from oligonucleotide 8334 (Fig. 5A) to oligonucleotide 8336 (Fig. 5B), the efficiency of initiating at the first AUG codon improved noticeably. This can be seen in three ways. (i) The synthesis of pre-CAT, which constituted only 15% of the polypeptide yield from SP64(8334)B38, increased to 50% with SP64(8336)B38. (ii) The ratio of pre-CAT to CAT, which was about 1:2 with SP64(8334)B34, was reversed with SP64 (8336)B34. A similar shift was seen when protein yields from the two B169 constructs were compared. (iii) Synthesis of the short form of CAT, which constituted about 10% of the polypeptide yield from SP64(8334)B171 (this was evident when the autoradiogram in Fig. 5A was exposed longer), was completely suppressed with SP64(8336)B171.

The only difference between the constructs tested in Fig. 5A and those used in Fig. 5B is the oligonucleotide adaptor that follows the first AUG codon. The sequence of oligonucleotide 8336 allows it to form a hairpin structure (horizontal arrows in Fig. 2A) that cannot form in oligonucleotide 8334. To determine whether the putative base-paired structure downstream from the first AUG codon accounts for the enhanced synthesis of pre-CAT by plasmids in series 8336, two additional constructs were made and tested. When three G residues in oligonucleotide 8336 were replaced by three C's [construct SP64(8414)B34 in Fig. 2C], the potential for base pairing was disrupted and the pre-CAT/CAT ratio

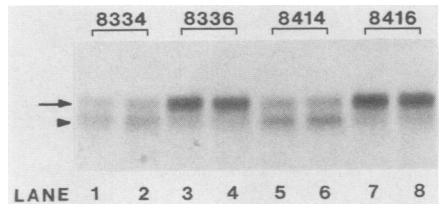


FIG. 6. Demonstration that downstream secondary structure increases initiation from the preceding AUG codon. The experiment was carried out with the reticulocyte translation system used for Fig. 4 and 5. The upstream AUG codon in all cases was from B34 (U in position −3 and G in position +4); the only difference among the mRNAs used was the sequence of the 22-nucleotide adaptor that followed the pre-CAT initiator codon. The sequences of the four adaptors, 8334, 8336, 8414, and 8416, are given in Fig. 2C. Adjacent lanes enclosed by brackets represent duplicate reactions carried out with each mRNA.

reverted to that obtained with oligonucleotide 8334 (Fig. 6, lanes 5 and 6 versus lanes 1 and 2). Preferential synthesis of pre-CAT was restored with oligonucleotide 8416 (Fig. 6, lanes 7 and 8), in which a second mutation restored the downstream hairpin structure (Fig. 2C, line 3).

Competition among mRNAs is not affected by sequences flanking the AUG codon. Because the scanning model (33) postulates that the AUG initiator codon serves not as a direct entry site for 40S ribosomal subunits but as a signal that halts 40S subunit migration, the model predicts that sequence variations around the AUG codon should not affect the ability of a message to compete for limiting 40S subunit-factor complexes. To test that prediction, I studied the translation of three SP64-derived transcripts in the presence and absence of excess rabbit β-globin mRNA (Fig. 5A, lanes 9 to 11). The inclusion of globin mRNA at a level that reduced overall CAT synthesis by about 75% caused no shift in the relative use of the first (weak) versus the second AUG codon in SP64(8334)B38 and -B34, nor was the translation of either construct more sensitive to competition than that of SP64(8334)B171.

In wheat germ extracts, context-dependent recognition of initiator codons depends on Mg^{2+} concentration. At 3 mM Mg^{2+}, wheat germ ribosomes did not respond properly to sequence changes near the AUG codon, and non-AUG codons were recognized far more efficiently than they should have been (Fig. 7A). When the same mRNAs were tested at 1.9 mM Mg^{2+}, however, the expected context effects were seen (Fig. 7B, lanes 1 to 5). At 1.9 mM Mg^{2+}, GUG and CUG, both in favorable contexts, were recognized with the appropriate low (≤10%) efficiency (Fig. 7B, lanes 7 and 8); there was no significant initiation at GUG when it occurred in a suboptimal context (Fig. 7B, lane 9) or at UUG even when it was in the optimal context (Fig. 7B, lane 6).

The improvement in the fidelity of initiation upon lowering the Mg^{2+} concentration (Fig. 7B versus 7A) was obtained at the cost of reduced amino acid incorporation, but that tradeoff is not inevitable. At 1.2 mM Mg^{2+} and 0.34 mM spermidine, overall amino acid incorporation was about as efficient as at 3 mM Mg^{2+}, whereas initiation retained a respectable dependence on context (Fig. 7C). The stimulatory effects of spermidine seem to be limited to the elongation or termination step in protein synthesis, since the formation of initiation complexes at 1.2 mM Mg^{2+}, monitored by glycerol gradient centrifugation, was equally efficient in the absence of spermidine as in its presence (data not shown).

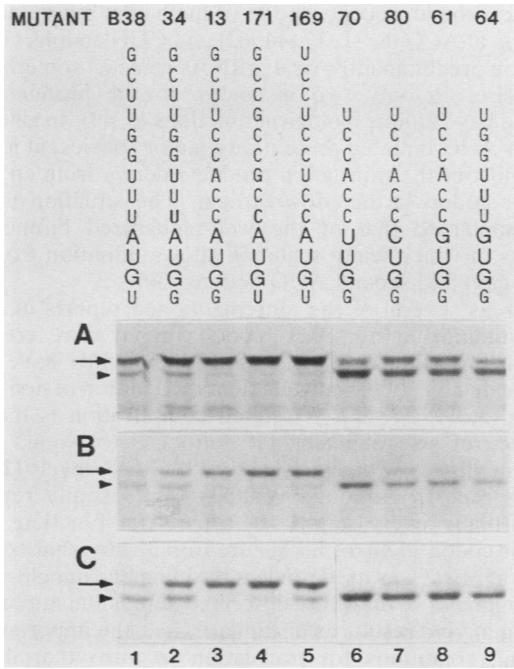


FIG. 7. In vitro translation of SP6 vector-derived transcripts in wheat germ extracts at 3 mM Mg^{2+} (A), 1.9 mM Mg^{2+} (B), or 1.2 mM Mg^{2+} plus 0.34 mM spermidine (C). The sequence around the pre-CAT start site is shown above each lane. All mRNAs contained oligonucleotide 8334 as the adaptor.

DISCUSSION

A reasonable approach to the use of in vitro translation systems is first to establish conditions under which cell-free systems reproduce what has been observed with whole cells; thus validated, in vitro systems may be used to explore novel questions about translation. Here I show that two characteristics of translation in vivo, sensitivity to sequences flanking the AUG codon and inefficient initiation at non-AUG codons, can be reproduced under appropriate conditions in vitro. Using the conditions thus defined, I show that context near the AUG codon is not a determinant of mRNA competition, consistent with the predictions of the scanning model (33). I also show, somewhat unexpectedly, that secondary structure downstream from a weak AUG codon can enhance recognition of the codon. The latter finding is the first evidence that secondary structure, which has long been recognized as a negative effector of translation (29, 34), can also affect initiation in a positive way.

Practical aspects of initiation in cell-free systems. With reticulocyte lysates, the optimal Mg^{2+} concentration for amino acid incorporation fortuitously coincides with the Mg^{2+} concentration that allows faithful initiation of translation. That is not true of wheat germ extracts. The overall incorporation of amino acids by wheat germ ribosomes is maximal at about 2.5 to 3.5 mM Mg^{2+} (20, 41, 53); however, that level of Mg^{2+}, which is used by most investigators, is clearly too high for faithful initiation (Fig. 7A). Supplementing the wheat germ system with 1.2 mM Mg^{2+} and 0.34 mM spermidine allows faithful initiation, i.e., context-dependent selection of the first versus the second AUG codon, without compromising the overall efficiency of amino acid incorporation.

There have been three previous reports of context effects on translation in vitro. One study found effects with ACG but not with AUG as the initiator codon (44). Another demonstrated context effects with reticulocyte but not wheat germ extracts (40). The third study (55) concluded that context effects in the reticulocyte system varied with Mg^{2+} concentration, but the authors were unable to define conditions under which the in vitro ratio of initiation from the first versus the second AUG codon mimicked the in vivo ratio. In contrast, this report describes conditions under which cell-free translation systems from wheat germ as well as rabbit reticulocytes show the same context requirements for recognition of AUG codons, and the same low efficiency of initiation at non-AUG codons, as occurs in vivo.

Although the reticulocyte translation system is at least as sensitive as the wheat germ system to sequence changes around the AUG codon, the reticulocyte system has other limitations that dictate caution in its use: reticulocyte lysates lose some activity during thawing and refreezing; the capacity for binding mRNA in an initiation assay is low in comparison with wheat germ extracts (M. Kozak, unpublished data); reticulocyte extracts are uniquely sensitive to inhibition by double-stranded RNA that commonly contaminates cellular mRNA preparations (51); and reticulocyte ribosomes, perhaps because they are less cap dependent than wheat germ ribosomes (39; Kozak, unpublished data), are prone to initiating translation on fragments of degraded mRNA. That tendency probably explains the internal initiation that is sometimes observed during experiments with reticulocyte lysates (see citations in reference 33). The internal initiation site detected when CAT mRNA was translated in HeLa cell extracts (46) was noticeably absent in the experiments described here (Fig. 5); it could be activated, however, by deliberately fragmenting the mRNA.

Theoretical aspects of initiation in eucaryotes. The in vitro data presented here confirm and extend two observations made previously in vivo. One is verification of the finding (28, 31) that a sequence encompassing 9 nucleotides upstream of the AUG codon, along with the G in position +4, facilitates recognition of the initiator codon. Although most of the evidence in that regard comes from experimentation with (28, 31) and compilation of (32) metazoan mRNA sequences, facilitating effects of A in position −3 and G in position +4 have also been demonstrated with plant systems in vivo (15, 60) and now in vitro (Fig. 7). The second point confirmed here is a specific prediction of the scanning model: that access to the second AUG codon will decrease as sequences flanking the 5′-proximal AUG codon increasingly resemble the consensus sequence for initiation. That point was established in vivo by varying the context around an upstream, out-of-frame AUG codon and showing that the yield of protein from a downstream site decreased as the conformity of the upstream site increased (28). In those experiments, however, I could only assume and could not directly measure initiation from the upstream AUG codon. The constructs used for the in vitro study described here were designed with the first and second AUG codons in the same reading frame, thereby enabling initiation from both AUG codons to be scored simultaneously. The results, in short, exactly conform to the predictions of the scanning model (33). The fact that translation of pre-CAT initiated from an AUG codon in a weak context was not preferentially inhibited under conditions of mRNA competition is also consistent with the scanning model, according to which recognition of the AUG codon is a late event that occurs after the competition-sensitive binding of a 40S ribosome-factor complex to the 5′ end of mRNA.

The most interesting new finding is that recognition of an

AUG codon in a suboptimal context improves considerably when a degree of secondary structure ($\Delta G = -19$ kcal/mol) is imposed just downstream. Thus, comparison of Fig. 5A and B shows a consistent reduction in CAT synthesis, and a corresponding increase in pre-CAT synthesis, when oligonucleotide 8334 was replaced by oligonucleotide 8336. The control constructs tested in Fig. 6 confirm that the potential for base pairing is what underlies the enhancing effect of oligonucleotide 8336. A reasonable explanation is that the predicted stem-loop structure in oligonucleotide 8336 causes the scanning 40S ribosome to pause, thereby providing more time for recognition of the preceding AUG codon. This is somewhat analogous to the idea that decoding by minor tRNAs might be assisted by strategically positioned stem-loop structures that cause ribosomes to pause during polypeptide elongation (58). In a different situation, a stem-loop structure near the *gag-pol* junction in Rous sarcoma virus mRNA has been shown to facilitate frameshifting (21), again perhaps by slowing the transit of ribosomes through that region.

The enhanced initiation attributed to the structure-prone 8336 insert in these experiments might be relevant to the translation of natural mRNAs in several ways. Although 97% of vertebrate mRNAs have a purine in position −3 (32), few possess the full consensus sequence; therefore, one might expect scanning to be slightly leaky in the majority of cellular mRNAs. That this is not the case might be attributed to the contribution of downstream secondary structure. That idea is supported by older experiments (24) in which chemical modification of reovirus mRNAs, in a way that ablated secondary structure, allowed 40S ribosomes to migrate beyond the normal 5′-proximal AUG codon. Conversely, the presence in some mRNAs of an extraordinarily G+C rich leader sequence (reviewed in reference 33) might be expected to encourage adventitious upstream initiation events. The recent finding that ribosomes initiate at upstream in-frame CUG codons in addition to the AUG triplet that starts the open reading frame of fibroblast growth factor (14, 50) might be attributed to slow transit of 40S ribosomal subunits across the very G+C rich leader sequence of that mRNA. Finally, there is a handful of vertebrate mRNAs in which the functional AUG initiator codon occurs in such a weak context (lacking the important purine in position −3 as well as the preferred G in position +4) that one may wonder how the initiation site is recognized at all. Six such mRNAs were identified in the most recent survey (32); one of them (38) was subsequently found to be an error (37); the sequences of the remaining five show a region of dyad symmetry just downstream from the AUG codon which might compensate, to a limited extent, for absence of the preferred primary sequence.

Initiation at non-AUG codons in eucaryotes. AUG appears to be the exclusive initiator codon in the hundreds of published mRNA sequences from vertebrate (32) and yeast (8) cells. An AUG-to-ACG mutation in the human α-globin gene reduces expression of the gene enough to cause clinical thalassemia (48); the expression of other cellular or viral genes is greatly reduced or abolished when the AUG initiator codon is converted to UUG (16, 52), ACG (7, 12, 61), AUC (54), or AUA (43, 56). For successful expression in mammalian cells, bacterial genes that initiate with GUG require the substitution of an AUG codon (35, 59), confirming the usual inability of eucaryotic ribosomes to initiate efficiently at non-AUG codons. Indeed, mutation of the initiator codon to GUG, UUG, or AUA appears to have been a step in the creation of some mammalian pseudogenes (2, 4, 22). Against that background, recent reports of inefficient but detectable initiation at ACG (6, 11, 17, 44, 62) and CUG codons (14, 19, 49, 50) in predominantly viral mRNAs came as something of a surprise. In only two instances do the proteins thus initiated have unique biological functions (6, 49). In all cases, the non-AUG-initiated proteins are minor species, at least in vivo; most of the translation products derive from an AUG initiator codon farther downstream. The situation is thus very similar to that of the well-recognized bifunctional mRNAs in which leaky scanning allows initiation from the first (weak) and second AUG codons (30).

Whereas several of the aforementioned reports of inefficient initiation at non-AUG codons in vivo are credible, attempts to explore the phenomenon in vitro have been complicated by abnormally efficient recognition of non-AUG codons (3, 17, 44, 65). A related complication is that the efficiency of selecting the first non-AUG (or weak AUG) codon in vitro is strongly influenced by Mg^{2+} levels (23, 25, 44; this report). Although Peabody (45) recently reported that virtually every non-AUG triplet can function as an initiator codon in vitro, his failure to indicate what concentration of Mg^{2+} was used renders the claim meaningless. The main emphasis of the study described here is the importance of using in vivo results as a standard, i.e., the importance of selecting conditions for translation in vitro that do not exaggerate the recognition of non-AUG codons beyond what normally occurs in vivo. Because Mg^{2+} levels do fluctuate in vivo (64), there might be some physiological condition under which initiation at non-AUG (or weak AUG) codons is more efficient than normal, but that possibility is purely hypothetical at present.

One artifact that can now be attributed to the inadvertent use of supraoptimal Mg^{2+} levels during in vitro translation is the tendency of certain viral mRNAs to form disomes in the presence of sparsomycin (1, 47, 63). The mRNAs in question have long 5′ noncoding sequences that include an AUU, ACG, or AUC triplet in a good context for initiation; there is no evidence that recognition of those upstream non-AUG codons, which occurs in wheat germ extracts at 4 mM Mg^{2+}, has any physiological relevance.

The fact that mammalian ribosomes initiate (albeit inefficiently) at non-AUG codons only when those codons occur in a favorable context (44; Fig. 3) suggests that contact with nearby nucleotides might compensate for a weakened codon-anticodon interaction. A similar principle evidently applies in procaryotes inasmuch as initiation at non-AUG codons is far more common in gram-positive than gram-negative bacteria and the Shine-Dalgarno sequence is, on the average, 2 or 3 nucleotides longer in mRNAs from gram-positive bacteria (18). Indeed, recent mutagenesis experiments in *E. coli* have directly demonstrated that initiation at UUG requires a stronger than normal Shine-Dalgarno interaction (66). The fact that context effects on initiation are less evident in yeasts than in higher eucaryotes (5, 9) might explain why yeast ribosomes do not normally initiate, even inefficiently, at non-AUG codons (10, 13).

ACKNOWLEDGMENTS

I am grateful to Lynne Hunter for constructing and testing the initial SP64 · CAT vector from which the plasmids used in this study were derived.

Research in my laboratory is supported by Public Health Service grant GM 33915 from the National Institutes of Health.

LITERATURE CITED

1. **Ahlquist, P., R. Dasgupta, D. S. Shih, D. Zimmern, and P. Kaesberg.** 1979. Two-step binding of eukaryotic ribosomes to

brome mosaic virus RNA 3. Nature 281:277–282.
2. **Andersen, R. D., B. W. Birren, S. J. Taplitz, and H. R. Herschman.** 1986. Rat metallothionein-1 structural gene and three pseudogenes, one of which contains 5'-regulatory sequences. Mol. Cell. Biol. 6:302–314.
3. **Anderson, C. W., and E. Buzash-Pollert.** 1985. Can ACG serve as an initiation codon for protein synthesis in eucaryotic cells? Mol. Cell. Biol. 5:3621–3624.
4. **Bachman, N. J., M. I. Lomax, and L. I. Grossman.** 1987. Two bovine genes for cytochrome c oxidase subunit IV: a processed pseudogene and an expressed gene. Gene 55:219–229.
5. **Baim, S. B., and F. Sherman.** 1988. mRNA structures influencing translation in the yeast *Saccharomyces cerevisiae*. Mol. Cell. Biol. 8:1591–1601.
6. **Becerra, S. P., J. A. Rose, M. Hardy, B. M. Baroudy, and C. W. Anderson.** 1985. Direct mapping of adeno-associated virus capsid proteins B and C: a possible ACG initiation codon. Proc. Natl. Acad. Sci. USA 82:7919–7923.
7. **Cattaneo, R., A. Schmid, D. Eschle, K. Baczko, V. ter Meulen, and M. A. Billeter.** 1988. Biased hypermutation and other genetic changes in defective measles viruses in human brain infections. Cell 55:255–265.
8. **Cigan, A. M., and T. F. Donahue.** 1987. Sequence and structural features associated with translational initiator regions in yeast—a review. Gene 59:1–18.
9. **Cigan, A. M., E. K. Pabich, and T. F. Donahue.** 1988. Mutational analysis of the *HIS4* translational initiator region in *Saccharomyces cerevisiae*. Mol. Cell. Biol. 8:2964–2975.
10. **Clements, J. M., T. M. Laz, and F. Sherman.** 1988. Efficiency of translation initiation by non-AUG codons in *Saccharomyces cerevisiae*. Mol. Cell. Biol. 8:4533–4536.
11. **Curran, J., and D. Kolakofsky.** 1988. Ribosomal initiation from an ACG codon in the Sendai virus P/C mRNA. EMBO J. 7:245–251.
12. **Dixon, L., J. Jiricny, and T. Hohn.** 1986. Oligonucleotide directed mutagenesis of cauliflower mosaic virus DNA using a repair-resistant nucleoside analogue: identification of an agnogene initiation codon. Gene 41:225–231.
13. **Donahue, T. F., and A. M. Cigan.** 1988. Genetic selection for mutations that reduce or abolish ribosomal recognition of the *HIS4* translational initiator region. Mol. Cell. Biol. 8:2955–2963.
14. **Florkiewicz, R. Z., and A. Sommer.** 1989. Human basic fibroblast growth factor gene encodes four polypeptides: three initiate translation from non-AUG codons. Proc. Natl. Acad. Sci. USA 86:3978–3981.
15. **Gallie, D. R., D. E. Sleat, J. W. Watts, P. C. Turner, and T. M. A. Wilson.** 1987. A comparison of eukaryotic viral 5'-leader sequences as enhancers of mRNA expression in vivo. Nucleic Acids Res. 15:8693–8711.
16. **Grass, D. S., and J. L. Manley.** 1987. Selective translation initiation on bicistronic simian virus 40 late mRNA. J. Virol. 61:2331–2335.
17. **Gupta, K. C., and S. Patwardhan.** 1988. ACG, the initiator codon for a Sendai virus protein. J. Biol. Chem. 263:8553–8556.
18. **Hager, P. W., and J. C. Rabinowitz.** 1985. Translational specificity in *Bacillus subtilis*, p. 1–32. *In* D. Dubnau (ed.), The molecular biology of the bacilli, vol. 2. Academic Press, Inc., New York.
19. **Hann, S. R., M. W. King, D. L. Bentley, C. W. Anderson, and R. N. Eisenman.** 1988. A non-AUG translational initiation in c-myc exon 1 generates an N-terminally distinct protein whose synthesis is disrupted in Burkitt's lymphomas. Cell. 52:185–195.
20. **Hunter, A. R., P. J. Farrell, R. J. Jackson, and T. Hunt.** 1977. The role of polyamines in cell-free protein synthesis in the wheat germ system. Eur. J. Biochem. 75:149–157.
21. **Jacks, T., H. D. Madhani, F. R. Masiarz, and H. E. Varmus.** 1988. Signals for ribosomal frameshifting in the Rous sarcoma virus gag-pol region. Cell 55:447–458.
22. **Kodaira, M., T. Kinashi, I. Umemura, F. Matsuda, T. Noma, Y. Ono, and T. Honjo.** 1986. Organization and evolution of variable region genes of the human immunoglobulin heavy chain. J. Mol. Biol. 190:529–541.
23. **Kozak, M.** 1979. Migration of 40S ribosomal subunits on messenger RNA when initiation is perturbed by lowering magnesium or adding drugs. J. Biol. Chem. 254:4731–4738.
24. **Kozak, M.** 1980. Influence of mRNA secondary structure on binding and migration of 40S ribosomal subunits. Cell 19:79–90.
25. **Kozak, M.** 1982. Analysis of ribosome binding sites from the s1 message of reovirus. J. Mol. Biol. 156:807–820.
26. **Kozak, M.** 1983. Translation of insulin-related polypeptides from messenger RNAs with tandemly reiterated copies of the ribosome binding site. Cell 34:971–978.
27. **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.
28. **Kozak, M.** 1986. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. Cell 44:283–292.
29. **Kozak, M.** 1986. Influences of mRNA secondary structure on initiation by eukaryotic ribosomes. Proc. Natl. Acad. Sci. USA 83:2850–2854.
30. **Kozak, M.** 1986. Bifunctional messenger RNAs in eukaryotes. Cell 47:481–483.
31. **Kozak, M.** 1987. At least six nucleotides preceding the AUG initiator codon enhance translation in mammalian cells. J. Mol. Biol. 196:947–950.
32. **Kozak, M.** 1987. An analysis of 5'-noncoding sequences from 699 vetebrate messenger RNAs. Nucleic Acids. Res. 15:8125–8148.
33. **Kozak, M.** 1989. The scanning model for translation: an update. J. Cell. Biol. 108:229–241.
34. **Kozak, M.** 1989. Circumstances and mechanisms of inhibition of translation by secondary structure in eucaryotic mRNAs. Mol. Cell. Biol. 9:5134–5142.
35. **Kwoh, T. J., D. Y. Kwoh, A. W. McCue, G. R. Davis, D. Patrick, and T. R. Gingeras.** 1986. Introduction and expression of the bacterial PaeR7 methylase gene in mammalian cells. Proc. Natl. Acad. Sci. USA 83:7713–7717.
36. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) 227:680–685.
37. **Landsman, D., O. W. McBride, N. Soares, M. P. Crippa, T. Srikantha, and M. Bustin.** 1989. Chromosomal protein HMG-14: identification, characterization, and chromosome localization of a functional gene from the large human multigene family. J. Biol. Chem. 264:3421–3427.
38. **Landsman, D., T. Srikantha, R. Westermann, and M. Bustin.** 1986. Chromosomal protein HMG-14: complete human cDNA sequence and evidence for a multigene family. J. Biol. Chem. 261:16082–16086.
39. **Lodish, H. F., and J. K. Rose.** 1977. Relative importance of 7-methylguanosine in ribosome binding and translation of VSV mRNA in wheat germ and reticulocyte cell-free systems. J. Biol. Chem. 252:1181–1188.
40. **Lütcke, H. A., K. C. Chow, F. S. Mickel, K. A. Moss, H. F. Kern, and G. A. Scheele.** 1987. Selection of AUG initiation codons differs in plants and animals. EMBO J. 6:43–48.
41. **Marcu, K., and B. Dudock.** 1974. Characterization of a highly efficient protein synthesizing system derived from commercial wheat germ. Nucleic Acids Res. 1:1385–1397.
42. **Melton, D. A., P. A. Krieg, M. R. Rebagliati, T. Maniatis, K. Zinn, and M. R. Green.** 1984. Efficient in vitro synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. Nucleic Acids Res. 12:7035–7056.
43. **O'Rear, J. J., and H. M. Temin.** 1982. Spontaneous changes in nucleotide sequence in proviruses of spleen necrosis virus, an avian retrovirus. Proc. Natl. Acad. Sci. USA 79:1230–1234.
44. **Peabody, D. S.** 1987. Translation initiation at an ACG triplet in mammalian cells. J. Biol. Chem. 262:11847–11851.
45. **Peabody, D. S.** 1989. Translation initiation at non-AUG triplets in mammalian cells. J. Biol. Chem. 264:5031–5035.
46. **Pelletier, J., M. E. Flynn, G. Kaplan, V. Racaniello, and N. Sonenberg.** 1988. Mutational analysis of upstream AUG codons of poliovirus RNA. J. Virol. 62:4486–4492.
47. **Pinck, M., C. Fritsch, M. Ravelonandro, C. Thivent, and L.**

Pinck. 1981. Binding of ribosomes to the 5′ leader sequence (N = 258) of RNA-3 from alfalfa mosaic virus. Nucleic Acids Res. **9**:1087–1100.
48. **Pirastu, M., G. Saglio, J. C. Chang, A. Cao, and Y. Wai Kan.** 1984. Initiation codon mutation as a cause of α thalassemia. J. Biol. Chem. **259**:12315–12317.
49. **Prats, A. C., G. De Billy, P. Wang, and J. L. Darlix.** 1989. CUG initiation codon used for the synthesis of a cell surface antigen coded by the murine leukemia virus. J. Mol. Biol. **205**:363–372.
50. **Prats, H., M. Kaghad, A. Prats, M. Klagsbrun, J. Lelias, P. Liauzun, P. Chalon, J. P. Tauber, F. Amalric, J. A. Smith, and D. Caput.** 1989. High molecular mass forms of basic fibroblast growth factor are initiated by alternative CUG codons. Proc. Natl. Acad. Sci. USA **86**:1836–1840.
51. **Pratt, G., A. Galpine, N. Sharp, S. Palmer, and M. J. Clemens.** 1988. Regulation of in vitro translation by double-stranded RNA in mammalian cell mRNA preparations. Nucleic Acids Res. **16**:3497–3510.
52. **Resnick, J., and T. Shenk.** 1986. Simian virus 40 agnoprotein facilitates normal nuclear location of the major capsid polypeptide and cell-to-cell spread of virus. J. Virol. **60**:1098–1106.
53. **Roberts, B. E., and B. M. Paterson.** 1973. Efficient translation of tobacco mosaic virus RNA and rabbit globin 9S RNA in a cell-free system from commercial wheat germ. Proc. Natl. Acad. Sci. USA **70**:2330–2334.
54. **Rosa, F., H. Berissi, J. Weissenbach, L. Maroteaux, M. Fellous, and M. Revel.** 1983. The β_2-microglobulin mRNA in human Daudi cells has a mutated initiation codon but is still inducible by interferon. EMBO J. **2**:239–243.
55. **Sangar, D. V., S. E. Newton, D. J. Rowlands, and B. E. Clarke.** 1987. All foot and mouth disease virus serotypes initiate protein synthesis at two separate AUGs. Nucleic Acids Res. **15**:3305–3315.
56. **Schlicht, H. J., G. Radziwill, and H. Schaller.** 1989. Synthesis and encapsidation of duck hepatitis B virus reverse transcriptase do not require formation of core-polymerase fusion proteins. Cell **56**:85–92.
57. **Shih, D. S., and P. Kaesberg.** 1976. Translation of the RNAs of brome mosaic virus: the monocistronic nature of RNA1 and RNA2. J. Mol. Biol. **103**:77–88.
58. **Shpaer, E. G.** 1985. The secondary structure of mRNAs from E. coli: its possible role in increasing the accuracy of translation. Nucleic Acids Res. **13**:275–288.
59. **Simonsen, C. C., M. Walter, and A. D. Levinson.** 1988. Expression of the plasmid-encoded type I dihydrofolate reductase gene in cultured mammalian cells: a novel selectable marker. Nucleic Acids Res. **16**:2235–2246.
60. **Taylor, J. L., J. D. G. Jones, S. Sandler, G. M. Mueller, J. Bedbrook, and P. Dunsmuir.** 1987. Optimizing the expression of chimeric genes in plant cells. Mol. Gen. Genet. **210**:572–577.
61. **Terwilliger, E., R. Burghoff, R. Sia, J. Sodroski, W. Haseltine, and C. Rosen.** 1988. The *art* gene product of human immunodeficiency virus is required for replication. J. Virol. **62**:655–658.
62. **Trempe, J. P., and B. J. Carter.** 1988. Alternate mRNA splicing is required for synthesis of adeno-associated virus VP1 capsid protein. J. Virol. **62**:3356–3363.
63. **Tyc, K., M. Konarska, H. J. Gross, and W. Filipowicz.** 1984. Multiple ribosome binding to the 5′-terminal leader sequence of tobacco mosaic virus RNA. Assembly of an 80S ribosome-mRNA complex at the AUU codon. Eur. J. Biochem. **140**:503–511.
64. **Walker, G. M.** 1986. Magnesium and cell cycle control: an update. Magnesium **5**:9–23.
65. **Wawrousek, E. F., and J. Piatigorsky.** 1987. Expression of the δ1-crystallin gene: salt-dependent alteration in the cell-free synthesis of δ-crystallin polypeptides. Gene **58**:155–165.
66. **Weyens, G., D. Charlier, M. Roovers, A. Pierard, and N. Glansdorff.** 1988. On the role of the Shine-Dalgarno sequence in determining the efficiency of translation initiation at a weak start codon in the *car* operon of *E. coli* K12. J. Mol. Biol. **204**:1045–1048.
67. **Zhang, H., R. Scholl, J. Browse, and C. Somerville.** 1988. Double stranded DNA sequencing as a choice for DNA sequencing. Nucleic Acids Res. **16**:1220.