JOURNAL OF VIROLOGY, Jan. 1990, p. 453–457
0022-538X/90/010453-05$02.00/0
Copyright © 1990, American Society for Microbiology
Vol. 64, No. 1

Translation Initiation at a Downstream AUG Occurs with Increased Efficiency When the Upstream AUG Is Located Very Close to the 5′ Cap

SYLVIA A. SEDMAN,† GREGORY W. GELEMBIUK, AND JANET E. MERTZ*

McArdle Laboratory for Cancer Research, University of Wisconsin, Madison, Wisconsin 53706

Received 19 June 1989/Accepted 4 October 1989

The major late 16S mRNA species of simian virus 40 encodes both a 61-amino-acid protein, LP1, and the major virion protein, VP1. Although the initiation signal GCCAUGG is usually utilized at high efficiency, at least one-third of 40S ribosomal subunits bypass it when it is present on the major 16S mRNA of simian virus 40 (S. A. Sedman, P. J. Good, and J. E. Mertz, J. Virol. 63:3884–3893, 1989). The LP1 translation initiation codon is situated 10 bases from the 5′ end of this mRNA. To determine whether the short length of the untranslated leader of this mRNA affects the efficiency of translation initiation at the LP1 initiation signal, monkey cells were transfected with plasmids which encode major late 16S-like mRNAs with 5′ untranslated regions (UTRs) of 6 or 44 bases. Decreasing the length of the 5′ UTR from 44 to 6 bases resulted in a 30% decrease in translation initiation at the LP1 AUG and a threefold increase in synthesis of VP1. When the VP1 open reading frame was replaced with the chloramphenicol acetyltransferase open reading frame, the reduction in 5′ UTR length resulted in a 70% decrease in translation initiation at the LP1 AUG and a 30% increase in chloramphenicol acetyltransferase synthesis. Therefore, ribosomes bypass an AUG codon more efficiently when it is located very close to the 5′ end of the mRNA.

Most capped eucaryotic mRNAs appear to be translated by a scanning mechanism (12). The initiator AUG codon is recognized on the basis of its position within the mRNA and its sequence context. According to the modified scanning model, the 40S ribosomal subunit binds at or near the 5′ end of the mRNA and scans toward the 3′ end until it encounters an AUG codon. If the AUG is situated in a favorable sequence context, the 40S ribosomal subunit stops, the 60S ribosomal subunit attaches, and translation begins. The optimal sequence context for efficient recognition as an initiation signal has been defined as $\overset{-3}{A}CC\overset{+1}{\underline{A}}\underline{UG}\overset{+4}{G}$, with the −3 and +4 positions being of primary importance (13, 14). The translation initiation signal is located between 20 and 100 bases from the 5′ end of most eucaryotic mRNAs (15). However, it is sometimes situated within 10 bases of the 5′ end (1, 3, 11, 17, 23). Results presented here show that 40S ribosomal subunits initiate translation less frequently at an initiation signal located 6 bases from the 5′ end of an mRNA than when it is located 44 bases from the 5′ end. Therefore, the efficiency of initiation of translation at an AUG codon depends upon its distance from the 5′ end of an mRNA as well as its sequence context.

The mRNAs used in the experiments presented here are variants of the major late 16S mRNA of simian virus 40 (SV40). This mRNA is functionally bigenic, encoding LP1, a 61-amino-acid protein (also called agnoprotein), and the major virion protein, VP1 (Fig. 1A). The AUG codon located closest to the 5′ end of this mRNA is the translation initiation signal used in the synthesis of LP1. Although it is present in a strong sequence context (GCCAUGG) for translation initiation, at least one-third of 40S ribosomal subunits bypass this initiation signal (19). Furthermore, when the optimal translation initiation sequence (ACCAUGG) is substituted for the LP1 initiation signal, at least 20% of ribosomes still bypass it (19). However, the major 16S mRNA of SV40 is unusual in that the LP1 initiation signal is located only 10 bases from the 5′ end (3). It is known that 40S ribosomal subunits protect approximately 60 bases of mRNA from digestion with nucleases (16). A plausible hypothesis to explain why ribosomes bypass the LP1 initiation signal at a high frequency is that the efficiency of initiation at an AUG codon is reduced when it is located within some minimal distance from the cap (12). To our knowledge, this hypothesis has never been directly tested in vivo.

To test the validity of this hypothesis, two plasmids, RSV(−6)LP1/VP1 and RSV(−44)LP1/VP1 (Fig. 1D), were constructed which, when transfected into COS cells, resulted in the synthesis of SV40 major late 16S mRNAs identical in sequence except for a difference in the lengths of their 5′ untranslated regions (UTRs). Because transcription initiates from the SV40 late promoter at numerous sites (7, 21), the Rous sarcoma virus (RSV) promoter was used to generate mRNAs with unique 5′ ends. The mRNAs expressed from the RSV promoter differed from the 16S mRNAs of wild-type SV40 in that they contained (i) a substitution of the sequence GGTCGACC in place of SV40 nucleotides (nt) 502 through 521, removing the LP1 termination codon, and (ii) a substitution of the sequence GGTCG in place of SV40 nt 1464 through 1492, removing two translation termination codons between the LP1 and VP1 open reading frames (ORFs). These alterations result in the fusion of the LP1 ORF to the VP1 ORF (Fig. 1B). The relative efficiencies of initiation of translation at both the LP1 and VP1 initiation codons could be measured since both VP1 and the LP1-VP1 fusion protein can be immunoprecipitated with polyclonal antiserum raised against VP1 (19). Therefore, we could measure the effect of 5′ UTR length on translation initiation at both the 5′ proximal AUG and an AUG located downstream. Ribosomes which initiate at the LP1 initiation signal are unable to reinitiate at the VP1 initiation signal.

* Corresponding author.
† Present address: Laboratory of Cellular Oncology, National Cancer Institute, Bethesda, MD 20892.

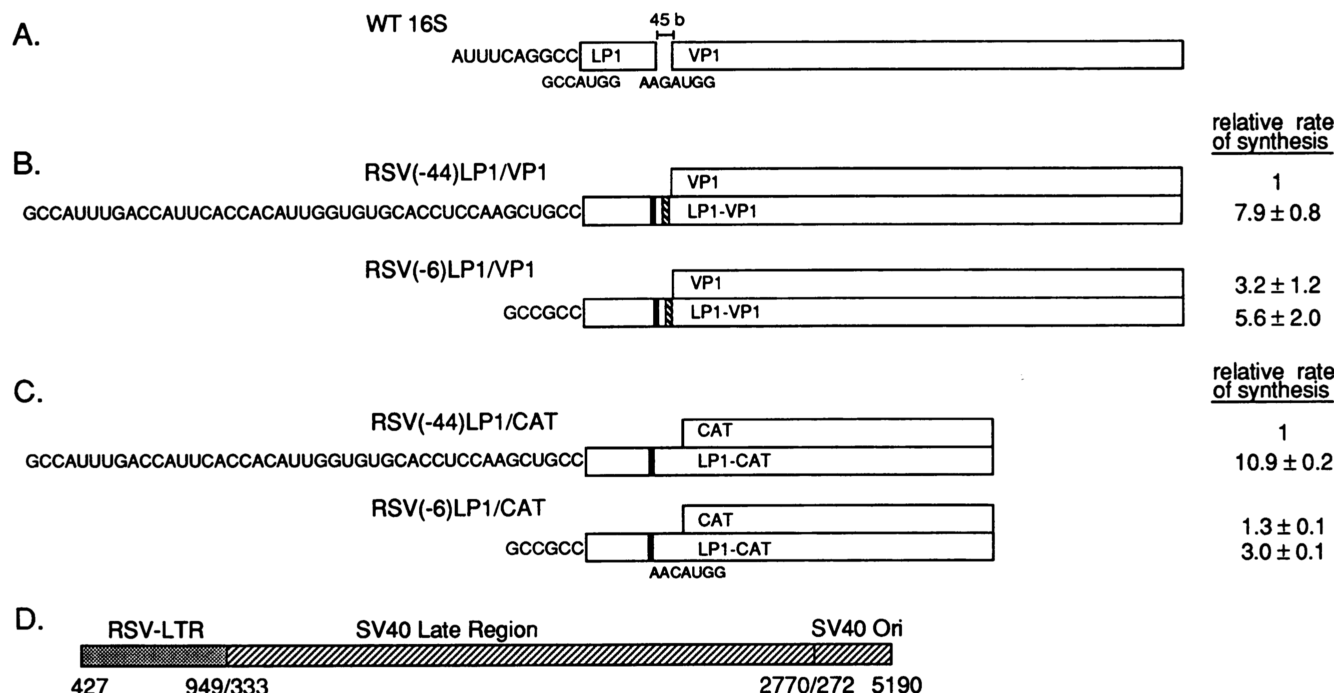

FIG. 1. ORF configurations on the major late 16S mRNA of SV40 (A) and the RSV promoter-driven VP1 (B) and CAT (C) mRNAs. Boxes represent the LP1, VP1, LP1-VP1, CAT, and LP1-CAT ORFs as indicated. LP1 is encoded by SV40 nt 335 through 520; VP1 is encoded by SV40 nt 1505 through 2590. To the left is shown the sequence of the 5′ UTR of each mRNA. Shown below the ORFs are the sequence contexts of the translation initiation signals. Vertical lines represent the sites of substitution and deletion mutations which result in the fusion of the LP1 and downstream ORFs. The sequences of RSV(−6)LP1/VP1 and RSV(−44)LP1/VP1 are identical except for the 5′ UTR. The same is true for RSV(−6)LP1/CAT and RSV(−44)LP1/CAT. The corresponding VP1- and CAT-encoding mRNAs are identical in sequence from the cap site to the beginning of the substitution mutation (▮) at SV40 nt 501. ▧, Substitution mutation at SV40 nt 1463. To the right of each ORF is a summary of the rate of synthesis of that protein relative to the rate of synthesis of VP1 from the RSV(−44)LP1/VP1 mRNA (B) and CAT from the RSV(−44)LP1/CAT mRNA (C). Relative rates of protein synthesis were determined by immunoprecipitation of pulse-labeled protein (see Fig. 3), with normalization to the amount of cytoplasmic SV40 late 16S-like mRNA present in cells transfected in parallel (see Fig. 4); they were not corrected for differences in protein half-lives. Listed are the means ± the standard errors of the means of data obtained from two independent experiments similar to the ones shown in Fig. 3 and 4. (D) Structure of pRSV(−44)LP1/VP1 after removal of the vector sequences. The plasmid consists of SV40 nt 346 through 2258 from pSV16S (fusion) (19), SV40 nt 2259 through 2770 from SV40 wild-type strain 800, and a fragment from pLC (fusion) (J. E. Mertz, G. Gelembiuk, and J. M. Kane, manuscript in preparation) which contains a minimal SV40 origin of replication (SV40 nt 5190 through 34 plus nt 197 through 272) and nt 427 through 949 of the RSV long terminal repeat (4) followed by linker DNA (CCAAGCUG) and SV40 nt 333 through 345. Symbols: ▩, sequences from RSV; ▨, sequences from SV40. pRSV(−6)LP1/VP1 is identical in sequence to pRSV(−44)LP1/VP1, except for the lack of 38 base pairs of sequence encoded in the 5′ UTR of the 16S mRNA (panel B). The SV40 ori sequence is adjacent to the RSV sequences in the plasmid. pRSV(−44)LP1/CAT is identical to pLC (fusion).

One additional AUG codon (not shown) is present upstream of the VP1 ORF; it is followed two codons thereafter by an in-frame translation termination signal.

The DNAs of these two plasmids were digested with *Bam*HI to excise the vector sequences, ligated into monomeric circles, and transfected into COS cells as described previously (20). Cells transfected with RSV(−6)LP1/VP1 were predicted to synthesize SV40 late mRNA containing a 6-base 5′ UTR, while cells transfected with RSV(−44)LP1/VP1 were predicted to synthesize SV40 late mRNA containing a 44-base 5′ UTR. Analysis by primer extension of the structures of the cytoplasmic mRNAs synthesized in cells transfected with these DNAs showed that greater than 95% of the viral mRNA made from each plasmid had unique 5′ UTRs of 6 and 44 bases, respectively (Fig. 2).

To determine the rate of synthesis of the LP1-VP1 fusion protein relative to that of VP1 for each mRNA, COS cells transfected with each DNA were incubated for 30 min with [^{35}S]methionine. Immediately thereafter, the cells were lysed, and the ratio of the two VP1-related proteins synthesized from each mRNA was determined by immunoprecipitation using antibodies raised against VP1 followed by electrophoresis, autoradiography, and densitometry. The rate of synthesis of the LP1-VP1 fusion protein relative to that of VP1 was fourfold higher in cells transfected with RSV(−44)LP1/VP1 than in cells transfected with RSV(−6)LP1/VP1 (Fig. 3A). Therefore, 40S ribosomal subunits initiate translation at the LP1 relative to the VP1 initiation signal more often on the mRNA with the longer 5′ UTR.

To compare the rates of synthesis of each protein from these two mRNAs, the amounts of LP1-VP1 fusion protein and VP1 were normalized to the amounts of RSV-driven 16S mRNA present in the cytoplasm of cells transfected in parallel. RNA levels were determined by quantitative S1 mapping as described previously (20) (Fig. 4A). The results obtained indicated that synthesis of the LP1-VP1 fusion protein decreased approximately 30%, while VP1 synthesis increased threefold in cells transfected with RSV(−6)LP1/VP1 compared with cells transfected with RSV(−44)LP1/VP1 (Fig. 1B). Therefore, scanning 40S ribosomal subunits bypass LP1 and initiate at the VP1 initiation signal more

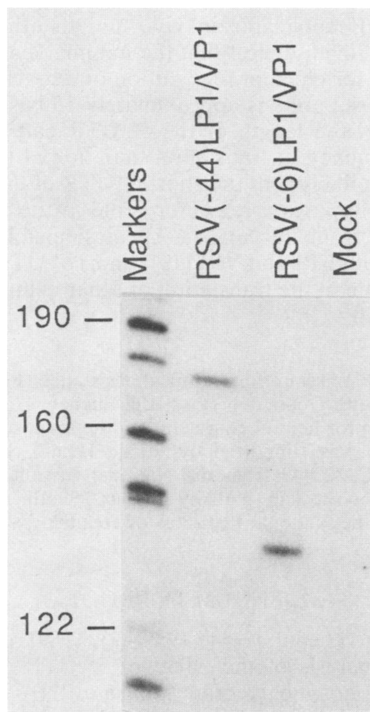

FIG. 2. Structure of the cytoplasmic SV40 late RNA made from RSV(−6)LP1/VP1 and RSV(−44)LP1/VP1. cDNAs of cytoplasmic RNA purified from 2 × 10^5 plasmid-transfected COS cells were synthesized by primer extension as described by Somasekhar and Mertz (21). The primer used was a 5′-end-labeled synthetic oligonucleotide spanning SV40 nt 470 through 453. The cDNAs were electrophoresed in a 5% polyacrylamide gel containing 8 M urea. Mock refers to cells that were transfected in parallel without DNA. Markers were 3′-end-labeled, *Msp*I-cut pBR322 DNA. Numbers on the left indicate sizes in nucleotides.

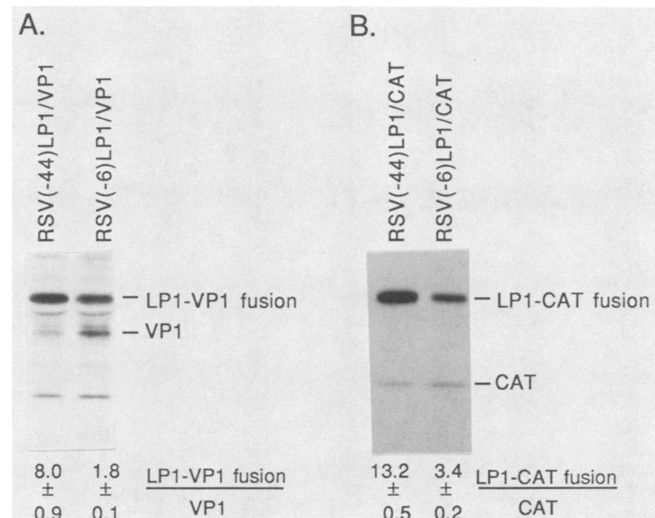

FIG. 3. Autoradiogram of pulse-labeled VP1-related (A) and CAT-related (B) proteins. The pBR322 vector sequences were excised from the plasmid DNAs by digestion of each plasmid with *Bam*HI. The viral sequences were then ligated to form monomeric circles before transfection into COS cells (1 μg of DNA per 10^6 cells) using DEAE-dextran as described by Sedman and Mertz (20). Forty hours later, cells were labeled for 30 min with [^{35}S]methionine and harvested immediately thereafter. Equal counts per minute were immunoprecipitated with polyclonal antiserum raised against VP1 (A) (20) and CAT (B) (5 Prime→3 Prime, Inc.) and analyzed by electrophoresis through 12% polyacrylamide gels containing sodium dodecyl sulfate as described previously (20). The gels were impregnated with fluor, dried, and exposed to X-ray film. The ratio of the two proteins synthesized from each mRNA, determined by densitometry, is indicated below each lane (mean ± standard error of the mean; data obtained from two experiments). Positions of the VP1, LP1-VP1 fusion, CAT, and LP1-CAT fusion bands are indicated. The half-lives of these proteins were determined to be >3, ~1, >5, and > 5 h, respectively (data not shown).

frequently when the LP1 AUG is located 6 bases from the 5′ end than when it is located 44 bases from the 5′ end.

To rule out the possibility that this finding was peculiar to the ORFs encoding the SV40 virion proteins, analogous constructs were analyzed in which the VP1 ORF was replaced by the ORF encoding chloramphenicol acetyltransferase (CAT) (Fig. 1C). S1 nuclease mapping indicated that only the RNA species depicted in Fig. 1C were made (data not shown). As with the LP1-VP1 constructs, the rate of synthesis of LP1-CAT relative to CAT was fourfold higher in cells transfected with RSV(−44)LP1/CAT compared with cells transfected with RSV(−6)LP1/CAT (Fig. 3B). Normalization of CAT and LP1-CAT fusion protein levels to levels of cytoplasmic CAT-encoding mRNA (Fig. 4B) indicated that synthesis of the LP1-CAT fusion protein decreased by 70% and that synthesis of CAT protein increased by 30% in cells transfected with RSV(−6)LP1/CAT compared with cells transfected with RSV(−44)LP1/CAT (Fig. 1C).

Therefore, for both pairs of mRNAs, changing the length of the 5′ UTR from 44 to 6 bases decreased the efficiency of initiation of translation at the LP1 initiation signal while increasing the efficiency of initiation at a downstream ORF. (The fact that the precise extent of the changes in the efficiencies of initiation at the two initiation signals differed between the two pairs of mRNAs was not the result of experimental error, since the standard errors in our experiments were small [Fig. 1]. Rather, the difference is probably a consequence of different downstream sequences having an effect on initiation at the 5′ proximal AUG codon [for similar observations, see references 2 and 22].) Are these changes in the efficiencies of initiation due to specific sequences present in the RSV-derived UTR? Some RNAs (e.g., tobacco mosaic virus RNA and alfalfa mosaic virus RNA 4) contain specific sequences within their 5′ UTR which significantly affect the overall efficiency with which the mRNA is translated (5, 6, 9). The 44-base 5′ UTR used in the present study contains 38 bases which are missing from the 6-base 5′ UTR. If this sequence encodes information that solely increases the overall efficiency of translation of the mRNAs, the rates of synthesis of the proteins encoded by the two ORFs present on each of the mRNAs should not have changed in opposite directions. Therefore, the presence of these additional 38 bases in the longer 5′ UTR must affect the frequency with which 40S ribosomal subunits bypass the LP1 initiation signal rather than simply the efficiency with which they bind to the 5′ end of the mRNA. In addition, Grass and Manley (8) found that the VP1 ORF was translated inefficiently when a 61-base sequence was added to the 5′ end of the SV40 major late 16S mRNA. These 61 bases are unrelated in sequence to the 5′ UTR present in RSV(−44)LP1/VP1. Therefore, the length of the 5′ UTR, rather than specific sequences, affects translation initiation at the LP1 ORF.

The precise relationship between the length of the 5′ UTR

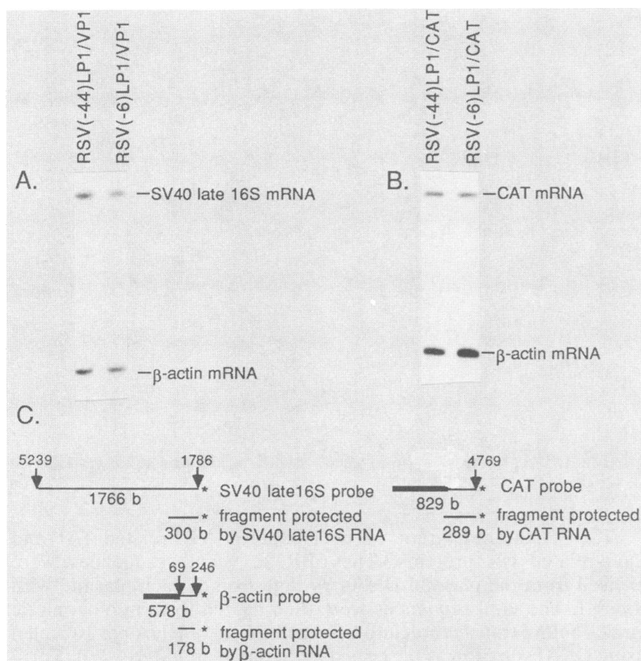

FIG. 4. Quantitative S1 mapping analysis. Cytoplasmic RNA samples harvested in parallel with the protein samples in Fig. 3 were S1 mapped simultaneously for cellular β-actin RNA and SV40 late 16S mRNA (A) and CAT mRNA (B). The ratio of these two bands in each sample was used as described previously (19) to normalize the relative amount of protein present in the corresponding sample in Fig. 3 to the amount of mRNA encoding that protein. This normalization corrects for possible differences in transfection efficiencies and half-lives of the mRNAs. Positions of the bands corresponding to the viral late 16S, CAT, and β-actin RNAs are indicated. (C) Strategy used for S1 nuclease mapping. The probe used to quantify SV40 late 16S mRNA, derived from pSV16S(−21) (19), contained SV40 nt 5239 through 1786 and was 5′ end labeled at nt 1786. Cellular β-actin mRNA was quantified by using a probe containing nt 69 through 246 of pHFβA-1 adjacent to 400 bases of vector DNA (19); it was 5′ end labeled at nt 246. Thin line represents actin sequences; thick line represents sequences derived from the vector. The probe used to quantify CAT mRNA contained 289 nt of homologous sequence adjacent to the unique *Eco*RI site (nt 4769) within the CAT gene plus 540 nt of nonhomologous sequence; it was 5′ end labeled at nt 4769. Thin line represents CAT sequences; thick line represents nonhomologous sequences.

and the frequency with which 40S ribosomal subunits bypass an initiation signal is not known, since our analysis only compared 5′ UTRs of two different lengths. Analysis of a series of plasmids which, when transfected into cells, result in the synthesis of SV40 major late 16S mRNA with diverse 5′ UTR lengths should enable us to answer this question. However, the minimal length of the 5′ UTR required for efficient initiation is probably at most 100 bases, since Johansen et al. (10) found that mRNAs with 5′ UTRs ranging from 101 to 184 bases were translated similarly.

Using an in vitro translation system, Strubin et al. (23) found that use of the 5′ proximal AUG of the Ia antigen-associated invariant chain mRNA increased relative to that of the downstream AUG when the 5′ UTR was 58 rather than 7 bases in length. Roner et al. (18) found that changing the length of the 5′ UTR of the reovirus s1 mRNA affected the overall efficiency of in vitro translation of this mRNA as well as the frequency of leaky scanning past the 5′ proximal AUG codon. If applicable in vivo and to other mRNAs, these latter findings suggest that the minimal length of the 5′ UTR required for efficient recognition of an AUG codon by 40S ribosomal subunits is approximately 15 bases.

In summary, the length of the 5′ UTR can significantly affect the frequency of translation initiation at the 5′ proximal AUG. It is likely that the short 5′ UTR of the major 16S mRNA of SV40 plays a regulatory role in the synthesis of LP1 and VP1 from this mRNA species, enabling efficient synthesis of both LP1 and VP1 (19). Short 5′ UTRs may play a regulatory role in the translation of other cellular and viral mRNAs as well.

We thank Bob Wisecup for technical assistance, Peter Good for plasmids, and Peter Good, Jeff Ross, Bill Sugden, Howard Temin, and Mike Young for helpful comments on the manuscript.

This research was supported by Public Health Service grants CA-07175 and CA-22443 from the National Institutes of Health. S.A.S. was supported in part by Public Health Service grant CA-09135 from the National Institutes of Health.

ADDENDUM IN PROOF

Van den Heuvel et al. (Gene **79**:83–95, 1989) recently have reported an analysis of the effect of 5′ UTR deletions on translation of phosphoglycerate kinase mRNA in *S. cerevisiae*. They found that mRNAs with 5′ UTR lengths of 27 or more bases were translated at wild-type efficiencies, whereas those with 5′ UTR lengths of 7 to 21 bases were translated at approximately half of this rate.

LITERATURE CITED

1. **Bos, J. L., L. J. Polder, R. Bernards, P. I. Schrier, P. J. van den Elsen, A. J. van der Eb, and H. van Ormondt.** 1981. The 2.2 kb E1b mRNA of human Ad12 and Ad5 codes for two tumor antigens starting at different AUG triplets. Cell **27**:121–131.
2. **Browning, K. S., S. R. Lax, J. Humphreys, J. M. Ravel, S. A. Jobling, and L. Gehrke.** 1988. Evidence that the 5′-untranslated leader of mRNA affects the requirement for wheat germ initiation factors 4A, 4F, and 4G. J. Biol. Chem. **263**:9630–9634.
3. **Celma, M. L., R. Dhar, J. Pan, and S. M. Weissman.** 1977. Comparison of the nucleotide sequence of the messenger RNA for the major structural protein of SV40 with the DNA sequence encoding the amino acids of the protein. Nucleic Acids Res. **4**:2549–2559.
4. **Czernilofsky, A. P., W. DeLorbe, R. Swanstrom, H. E. Varmus, J. M. Bishop, E. Tischer, and H. M. Goodman.** 1980. The nucleotide sequence of an untranslated but conserved domain at the 3′ end of the avian sarcoma virus genome. Nucleic Acids Res. **8**:2967–2984.
5. **Gallie, D. R., D. E. Sleat, J. W. Watts, P. C. Turner, and T. M. A. Wilson.** 1987. The 5′-leader sequence of tobacco mosaic virus RNA enhances the expression of foreign gene transcripts *in vitro* and *in vivo*. Nucleic Acids Res. **15**:3257–3273.
6. **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.
7. **Ghosh, P. K., V. B. Reddy, J. Swinscoe, P. Lebowitz, and S. M. Weissman.** 1978. Heterogeneity and 5′-terminal structures of the late RNAs of simian virus 40. J. Mol. Biol. **126**:813–846.
8. **Grass, D. S., and J. L. Manley.** 1987. Selective translation initiation on bicistronic simian virus 40 late mRNA. J. Virol. **61**:2331–2335.
9. **Jobling, S. A., and L. Gehrke.** 1987. Enhanced translation of chimaeric messenger RNAs containing a plant viral untranslated leader sequence. Nature (London) **325**:622–625.
10. **Johansen, H., D. Schumperli, and M. Rosenberg.** 1984. Affecting

gene expression by altering the length and sequence of the 5′ leader. Proc. Natl. Acad. Sci. USA **81**:7698–7702.

11. **Kelley, D. E., C. Coleclough, and R. P. Perry.** 1982. Functional significance and evolutionary development of the 5′ terminal regions of immunoglobulin variable region genes. Cell **29**:681–689.
12. **Kozak, M.** 1981. Mechanism of mRNA recognition by eukaryotic ribosomes during initiation of protein synthesis. Curr. Top. Microbiol. Immunol. **93**:81–123.
13. **Kozak, M.** 1984. Compilation and analysis of sequences upstream from the translational start sites in eukaryotic mRNAs. Nucleic Acids Res. **12**:857–872.
14. **Kozak, M.** 1986. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. Cell **44**:283–292.
15. **Kozak, M.** 1987. An analysis of 5′-noncoding sequences from 699 vertebrate messenger RNAs. Nucleic Acids Res. **15**:8125–8148.
16. **Kozak, M., and A. J. Shatkin.** 1978. Migration of 40 S ribosomal subunits on messenger RNA in the presence of edeine. J. Biol. Chem. **253**:6568–6577.
17. **Peterson, C. A., and J. Piatigorsky.** 1986. Preferential conservation of the globular domains of the βA3/A1-crystallin polypeptide of the chicken eye lens. Gene **45**:139–147.
18. **Roner, M. R., R. K. Gaillard, Jr., and W. K. Joklik.** 1989. Control of reovirus messenger RNA translation efficiency by the regions upstream of initiation codons. Virology **168**:292–301.
19. **Sedman, S. A., P. J. Good, and J. E. Mertz.** 1989. Leader-encoded open reading frames modulate both the absolute and relative rates of synthesis of the virion proteins of simian virus 40. J. Virol. **63**:3884–3893.
20. **Sedman, S. A., and J. E. Mertz.** 1988. Mechanisms of synthesis of virion proteins from the functionally bigenic late mRNAs of simian virus 40. J. Virol. **62**:954–961.
21. **Somasekhar, M. B., and J. E. Mertz.** 1985. Sequences involved in determining the locations of the 5′ ends of the late RNAs of simian virus 40. J. Virol. **56**:1002–1013.
22. **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.
23. **Strubin, M., E. O. Long, and B. Mach.** 1986. Two forms of the Ia antigen-associated invariant chain result from alternative initiations at two in-phase AUGs. Cell **47**:619–625.